ACTION DESIGN RESEARCH FOR APPLIED VISUALIZATION DESIGN

by

Nina McCurdy

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computing

School of Computing

The University of Utah

May 2020

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Nina McCurdy**

has been approved by the following supervisory committee members:

Miriah Meyer,	Chair(s)	**13 Sep 2019** Date Approved
Jason Dykes,	Member	**13 Sep 2019** Date Approved
Charles Hansen,	Member	**12 Sep 2019** Date Approved
Alexander Lex,	Member	**13 Sep 2019** Date Approved
Jason Wiese,	Member	**23 Sep 2019** Date Approved

by **Ross Whitaker**, Chair/Dean of

the Department/College/School of **Computing**

and by **David Kieda**, Dean of The Graduate School.

ABSTRACT

In applied visualization research, artifacts are shaped by a series of small design decisions, many of which are evaluated quickly and informally via methods that often go unreported and unverified. Such design decisions are influenced not only by visualization theory, but also by the people and context of the research. While existing applied visualization models support a level of reliability throughout the design process, they fail to explicitly address the influence of the research context in shaping the resulting design artifacts. In this work, we look to *action design research* (ADR) for insight into this gap. In particular, ADR offers a framework along with a set of guiding principles for navigating and capitalizing on the disruptive, subjective, human-centered nature of applied design research, while aiming to ensure the reliability of the process and design. This dissertation explores the utility of ADR for applied visualization research. Our exploration is grounded in a formative design study with poetry scholars, informed by preliminary theoretical research into the ADR framework, and developed in two consecutive design studies — the first in collaboration with global health experts, and the second in collaboration with astronomers and astrophysicists. Our exploratory results validate ADR as a useful model for strengthening the visualization research process, while also revealing significant gaps that pose important areas for future visualization research. Primary contributions of this dissertation include an articulation of the gaps in existing visualization methodology, an exploration of action design research for applied visualization design, and a reflective synthesis of the exploratory results. Secondary contributions stem from the results of the three design studies and reflect the development of our research thinking around artifacts, the emergent design process, and the generation of visualization knowledge.

To Coby, Nico, and Yoko

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Miriah Meyer, for guiding and supporting me through every step of this process. Working with Miriah has been such an enjoyable and rewarding experience. I feel incredibly fortunate to have had her as my advisor. I would also like to express my appreciation to my supervisory committee for their support and guidance in navigating this dissertation: Jason Dykes, Charles Hansen, Alex Lex, and Jason Wiese. It has been an honor working with them.

I am deeply grateful to the many people who contributed to this dissertation as coauthors, collaborators, and colleagues: Katharine Coles, Julie Lein, Craig Dworkin, Amanda Hurtado, Michael Gleicher, Vivek Srikumar, Jules Penham, Jason Dykes, Julie Gerdes, Mike Kirby, Alexander Lex, Erik Brunvand, Joel Primack, Sandra Faber, David Koo, Clayton Strawn, Vivian Tang, Chris Henze, Pat Moran, Tim Sandstrom, Bron Nelson, and David Ellsworth.

I would like to thank the current and past members of the Vis Design Lab at the University of Utah for their feedback, support, and for helping make these last six years such an enjoyable experience: Alex Bigelow, Aspen Hopkins, Sean McKenna, Ethan Kerzner, Jimmy Moore, Carolina Nobre, Jennifer Rogers, Sam Quinan, and Kris Zygmunt, Ilkin Safarli, Kiran Gadhave, Haihan Lin, Dylan Wooten, and Pascal Goffin. Special thanks (also) to Christine Pickett for her tremendous writing help throughout this process.

I would also like to thank my community: to Anne, Jonathan, and Coby, for their continuous support, encouragement, and feedback; to Hannah, Adrià, Lluc and Jaël, for their support from across the Atlantic; to my extended family and friends for keeping me sane and remaining interested in my research; and finally, to Nico, for giving me time to wrap things up.

This dissertation extends published work for which we have permission to reuse: Reprinted [1], with permission, McCurdy, Nina and Lein, Julie, and Coles, Katharine, and Meyer, Miriah; "Poemage: Visualizing the sonic topology of a poem;" *Transactions on Visualization and Computer Graphics;* ©2016 IEEE. Reprinted [2], with permission, McCurdy, Nina and Dykes, Jason, and Meyer, Miriah; "Action design research and visualization design;" *in Proceedings of the Sixth*

Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization; ©2016 ACM. Reprinted [3], with permission, McCurdy, Nina and Gerdes, Julie and Meyer, Miriah; "A framework for externalizing implicit error using visualization;" *Transactions on Visualization and Computer Graphics;* ©2018 IEEE. Reprinted [4], with permission, McCurdy, Nina and Meyer, Miriah; "GalStamps: Analyzing real and simulated galaxy observations;" *in Proceedings of the 2019 IEEE Visualization Conference;* ©2019 IEEE.

This dissertation was supported in part by NSF grant IIS-1350896, NEH grant HD-229002, and NSF grant GRF-1747505. Views expressed in this dissertation do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CHAPTER 1

INTRODUCTION

Applied visualization research is contextual, subjective, emergent, and disruptive. The design and development of technology artifacts is guided by visualization theory, driven by insight and experimentation, and heavily influenced by the research environment and by the knowledge and perspectives of the research collaborators. Existing applied visualization models promote the reliability of technology artifacts through grounding in visualization theory and validation within the application domain. These models, however, fail to account for the shaping of technology artifacts by the research context and members of the collaboration, and fail to provide explicit guidance on how to approach this shaping in a way that is reliable and also leverages the learning that occurs. Stemming from similar challenges within information systems research, a recently proposed methodology called action design research (ADR) offers potential insight into resolving this gap [5]. In this dissertation, we explore the utility of ADR for navigating and leveraging the contextual, human-centered nature of applied visualization research, and for increasing the reliability of technology artifacts. Our exploration is motivated by a formative design study [1], guided by preliminary theoretical research [2], and applied and developed in two design studies [3], [4]. Our results validate ADR as a useful model for strengthening the visualization design process, while also revealing important areas for future visualization methodology research.

1.1 Overview

Throughout the visualization design process, technology artifacts are shaped by a series of design decisions, only some of which are formally validated. Many of these decisions are instead evaluated through quick, informal, and lightweight mechanisms, the majority of which go unreported and unverified. In collaborative settings, these design decisions are influenced not just by visualization theory and guidelines, but also by the people and context in which the artifacts are designed. Furthermore, the design process is characteristically wicked [6] and emergent, driven by an evolving understanding of the data and tasks and the problem being addressed.

Documented adherence to applied visualization process [7], [8] and decision [9], [10] models affords a level of reliability for the resulting technology artifacts. These models stress reliability through the grounding of design decisions in established visualization principles and the validation of artifacts within the application domain. They do not, however, explicitly address the more subjective shaping of artifacts by the people and context involved in the project — a shaping that can at times ignore or even go against established visualization conventions [1]. They additionally do not explicate the role that deliberate disruption on the part of the visualization designer plays in the shaping of artifacts. Taken together, these gaps reveal important influencing factors within the visualization design process for which there are not yet established guidelines for ensuring the reliability of the resulting visualization technology artifacts.

Appropriately capturing and reporting on design decisions and their influencing factors in real time throughout the design process has the potential to greatly increase reliability, and recent work explores approaches to doing this effectively in applied visualization research contexts [11], [12]. Such work emphasizes the important role of concurrent, ongoing, and documented reflection. Current visualization models, however, largely defer efforts to reflect and report to the final stages of research when results are prepared and disseminated [13]. Our experience suggests that this approach is overly reliant on memory and subject to confirmation bias, and fails to leverage the learning that results from these activities in the design process itself.

Within information systems research, a recently proposed methodology called *action design research* (ADR) offers a framework that aims to increase the reliability of technology artifacts through adherence to a set of principles [5]. Like visualization design research, ADR seeks to contribute design knowledge by solving real-world problems, while supporting the messy, iterative, human-centered nature of the design process. ADR explicitly incorporates approaches from social science that acknowledge and facilitate the effects of people and context on the shaping of designed solutions, and specifically those that occur when actions taken by the researcher result in a disruption of the target users' processes or understanding, and vice versa [14]. This explicit incorporation of established social science approaches, namely those from action research [15], [16], provides a framework for recognizing and articulating both the role of disruption and the influence of people and context in and on the emergent design process and decision-making. Lacking from ADR, however, is explicit guidance on how to leverage this framework in order to increase reliability through appropriately capturing and reporting on design decisions. We further investigate this and other weaknesses of ADR as part of a larger reflective synthesis of the results of this dissertation research.

This dissertation explores the utility of ADR for navigating and leveraging the disruptive, subjective, contextual, and emergent nature of applied visualization research, and for increasing the reliability of visualization technology artifacts. The exploration is motivated by a formative design study, guided by preliminary theoretical research on the topic, and further investigated in two design studies. The formative design study, a 2-year collaboration with poetry scholars [1], formed the basis of our understanding of the disruptive, subjective, emergent design process and the span of learning that occurs, as well as our understanding of the gaps in existing visualization methodology. The theoretical research, examining ADR and its application to applied visualization research [2], provided us with an in-depth understanding of the ADR framework, as well as a preliminary understanding of how ADR concepts could be adapted and extended to better support the visualization design process. This work additionally helped us formalize our thinking about visualization research artifacts, the shaping of the design process, and the development of a broader range of visualization knowledge. Results from the theoretical research were further explored and developed over the course of two design studies — an 18-month collaboration with global health experts [3] and a 1-year collaboration with astrophysicists [4]. For each of these projects, we critically reflect on our use of ADR, synthesizing our results to: 1) provide evidence that ADR strengthens visualization design research by more adequately supporting the contextual, human-centered nature of the visualization design process; and 2) highlight significant gaps in the ADR framework and discuss important areas of future work in visualization methodology research.

This dissertation focuses on technology artifacts; however, our exploration suggests a broader class of research artifacts that are important for increasing both the reliability and transferability of applied visualization research. We discuss this further in a final reflective synthesis of our exploratory results.

1.2 Contributions

This dissertation explores the utility of action design research for navigating and leveraging the disruptive, subjective, contextual, and emergent nature of applied visualization research. Primary contributions include an articulation of the gaps in existing visualization methodology, grounded in a formative design study; an exploration of action design research for applied visualization research, grounded in theory and applied and developed in practice; and a reflective synthesis of exploratory results, articulating and discussing the elements of ADR that strengthen the visualization design process, as well as important elements that the framework fails to adequately address.

Secondary contributions of this dissertation stem from the results of the three design studies. In the design study with poetry scholars [1], the research contributions include a problem characterization and data abstraction for visualizing sonic devices in poetry; a validated instantiation of the characterization and abstraction in a tool for visualizing the sonic topology of a poem; and a reflection on the challenges of conducting applied visualization research in the domain of poetry scholarship. Additional research contributions include a formalism for analyzing sonic devices in poetry and an implementation of the formalism in a tool called RhymeDesign [17], and a 3D extension of Poemage exploring the visualization of sonic depth in poetic texts [18], [19]. In the design study with global health experts [3], research contributions include a framework for externalizing implicit error; an instantiation of the framework in an interactive visualization system designed to support the externalization of implicit error in Zika epidemiological data; a rich, reflective description of the research process; and preliminary prototyping of mechanisms to support recording, reflecting, and reporting in design studies. In the design study with astronomers and astrophysicists [4], research contributions include a data and task abstraction for statistically and visually analyzing real and simulated galaxy observations, as well as an initial design, implemented in a prototype called GalStamps, and evaluated through two case studies with domain experts. These secondary contributions reflect the development of our research thinking around artifacts, the emergent design process, and the generation of visualization knowledge.

1.3 Organization

The remainder of the dissertation is organized as follows: In Chapter 2, we present the relevant background on visualization design study, including the challenges surrounding reliability and the associated gaps in existing visualization models. We then present action design research (ADR) in the context of applied visualization research. In Chapter 3, we present the details of our formative design study with poetry scholars and discuss the ways in which the study motivated our exploration of ADR. In Chapter 4, we present and critique the poetry design study by reframing it through the lens of ADR. In Chapters 5 and 6, we present the details of our two consecutive design studies with global heath experts and with astronomers and astrophysicists, and reflect on our experience applying and extending ADR within each of these contexts. In Chapter 7 we present a reflective synthesis of our exploratory results. We conclude, in Chapter 8, with a summary of our process and results. This dissertation excerpts and extends previously and soon-to-be published work [1]–[4].

CHAPTER 2

BACKGROUND AND RELATED WORK

This dissertation investigates the contextual, human-centered nature of applied visualization research and the degree to which action design research (ADR) can help to address existing threats to the reliability of the research process and resulting technology artifacts. In this dissertation, design studies function as the primary method for conducting applied visualization research. In this chapter, we discuss the nature of design study research, the consequential threats to reliability and generalizability, and the associated gaps in current applied visualization methodology. We then present ADR in the context of applied visualization research by describing the framework in the language and constructs developed within the visualization community. Finally, we compare ADR to existing visualization methodology.

2.1 Design Study

We begin this section by defining several key characteristics of visualization design studies:

- **Contextual**: We use the term *contextual* to describe design processes and solutions that are highly specific; situated in and inseparable from the design research context including the members and dynamics of the research team and the environment in which the research was carried out [14], [20]–[22].
- **Disruptive**: Our notion of *disruptive* design research stems from action research in which disruption and intervention form the basis for learning about a problem [14], [23]. Disruption refers to anything that alters the way experts perceive of or interact with their domain, from reinterpreting domain concepts to introducing new technologies.
- **Emergent**: Borrowing from information systems research, the term *emergent* describes design processes and solutions that are informed by theory but guided by insights that arise through interactions with the problem domain, collaborators, and the research environment [5].
- **Subjective**: We use the term *subjective* to describe design processes and solutions that are influenced by the individual backgrounds, perceptions, values, interests, and assumptions of

all members of the research team, domain experts and visualization researchers alike [5], [21].

- **Wicked**: Rooted in the design community, the term *wicked* describes a class of design problems that are ill-defined, indeterminate, and marked by large solution spaces, no stopping points, and no ground truths. Solutions to wicked problems are unique and are neither right nor wrong, only good, bad, or somewhere in between [6], [24].

Visualization design studies aim to derive new visualization knowledge through developing effective solutions to real-world problems. In a design study, visualization researchers collaborate closely with domain experts to develop tools to support the specific domain analysis, while simultaneously using those contexts to identify and contribute to open visualization questions [8]. Thus, design study researchers are both researchers and practitioners. New visualization knowledge is derived from the learning that occurs throughout the design study and that shapes the design process and resulting artifacts [5]. Importantly, we distinguish visualization design study, defined above and referred to throughout this dissertation as *design study*, from the academic discipline of design study focused on developing a critical understanding of design processes across application domains.

Design studies offer a rich, ecologically valid environment for conducting applied visualization research; however, the nature of design study poses a threat to conventional notions of reliability and generalizability. Design studies are messy, iterative, and emergent [7], [8]; driven by discovery, insight, and multiple lines of inquiry; and rooted in the disruption of domain thinking and practice. They are semistructured and semi-improvised, and they require flexibility of process and methodology. The design study process is inherently difficult to capture and communicate, leaving much of the study closed to scrutiny, and requiring readers to place their faith in the researchers' actions and abilities [12], [25]. Additionally, design studies are highly contextual and subjective, situated within a specific environment and heavily influenced by this environment and by the knowledge and experience of participating members [5]. They often result in very specific systems. Whether these systems generalize to other visualization research contexts is often unclear. Design study has become a valued approach to applied visualization research, but these challenges prompt questions surrounding the reliability and generalizability of the design study process and contributions [26].

We use the word *reliability* to describe the extent to which the design process and contributions can be trusted — that is, that the design process occurred in the way it was reported and that claims surrounding resulting contributions are valid. Existing applied visualization models support a level of reliability [7]–[10] through adherence to established visualization principles and validation techniques. This approach, however, fails to address the critical role of people and environment in shaping the research process and results. Researchers in other fields specializing in qualitative and field-based research have thought deeply about issues of reliability in dynamic, contextual, and subjective research settings. Within the field of qualitative social science, the reliability of research is established in part by systematically documenting research practices and the emergence of findings and by allowing such documentation to be audited by other researchers [21]. Recent work within the visualization community explores approaches to enhancing documentation in design research [11], [12], [25] and to presenting auditable results [11]. This work marks the beginning of an important discussion about what reliability in design study looks like, and how it can be established and supported throughout the design process.

Once reliability has been established and researchers have gained the necessary trust in a design study's process and contributions, *transferability* measures the extent to which aspects of the process and contributions can be transferred or adapted to other research contexts. Transferability is already recognized as an important objective in design study. In the nine-stage framework for conducting design studies, transferability is cited as the more suitable alternative to reproducibility [8]. More recently, transferability has been reconfirmed as a common goal for all design study contributions [26]. According to these works, which draw from ethnography, establishing transferability entails sharing "sufficient knowledge about a solution that it may potentially be transferred to other contexts" [8]. Accordingly, the problem characterization and abstraction, an established contribution of design studies, may be cast as a way to promote transferability of findings [8]. Promoting transferability beyond the data and task abstraction, however, is difficult given the complex nature of design study research, and existing visualization methodology offers little guidance on how to do this effectively.

The concept of transferability is fundamental not just to ethnography but to many areas of qualitative, subjective, field-based research. It is widely considered the more suitable alternative to generalizability in qualitative and subjective research contexts [21]. This shift in terminology emphasizes the unique, contextual nature of individual studies. It also removes the unrealistic requirement that researchers have enough of an understanding of all receiving contexts, such that they can make precise statements about how their findings will apply more broadly [21]. Similar to the design study literature, the task of supporting transferability entails supplying readers with a database of contextual details that enables transferability judgments by potential appliers. This is accomplished primarily through appropriately thick description [21], a concept discussed further in Chapter 7.

Derived from a branch of the social sciences, namely action research [15], [16], action design research [5] is an attempt to incorporate established methods for increasing reliability in qualitative research into the design research process. In the next section, we take a critical look at action design research and discuss its relationship to, and implications for, applied visualization research.

2.2 Action Design Research

Developed within information systems research, action design research (ADR) [5] seeks to account for and leverage the role of people and context in shaping the design process and the development of technology artifacts. Through a set of guiding principles and a high-level process model, ADR provides a framework for navigating the contextual, human-centered nature of applied design research, and offers potential insight into how we can begin to address gaps in current visualization methodology. ADR has been reported as a development methodology in visual analytics systems research [27], but it has yet to be explored in light of applied visualization design methods. In the following sections, we discuss the development of the methodology within the information systems community, followed by an explication of each of the principles and process stages. Each discussion includes an analysis of how ADR concepts reflect, apply to, or inform current visualization design research theory and practice.

2.2.1 Overview

Information systems research is driven by a dual mission: to generate valuable information systems knowledge and to create effective solutions to real-world problems that inform an application's domain [5]. Early process models to achieve this mission focus on incorporating both experimental and design methods, while emphasizing relevance through a grounding in real-world problems [22], [28]. Sein *et al.* [5] critique this approach for not recognizing or capturing the influence of people, organizations, and context on shaping technology throughout the design process. They argue further that appropriate forms of evaluation that consider such elements must be interwoven throughout the design process.

The ADR methodology [5], shown in Fig. 2.1 (adapted [5], Fig. 1), is an attempt to address these shortcomings. Specifically, ADR explicates the influence of the research environment as well as the

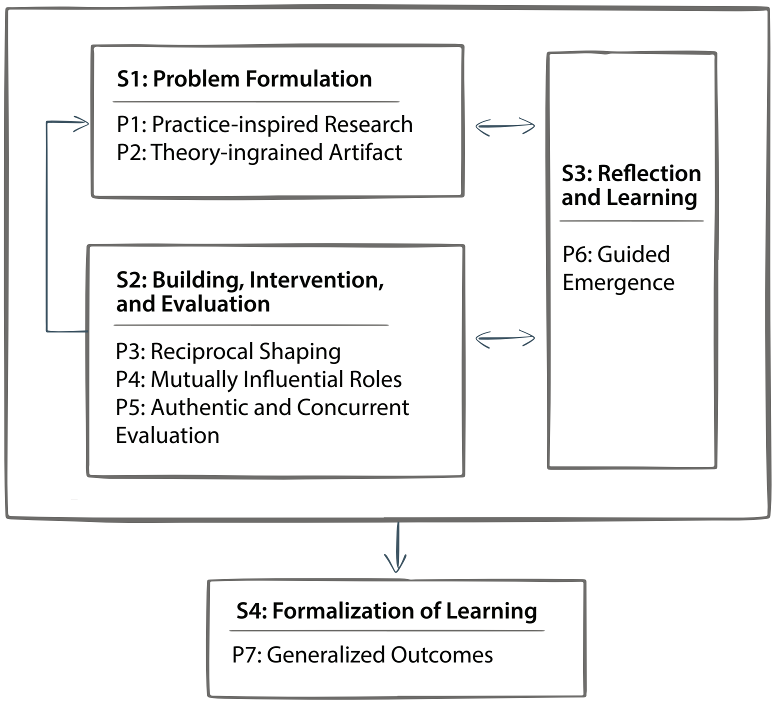

Fig. 2.1. Overview of the ADR methodology (adapted from [5], Fig. 1.) indicating the relationship of the stages (S1-S4) to each other, as well as the guiding principles (P1-P7). The deliberate omission of an arrow pointing from stage 1 to stage 2 may seem counterintuitive, but it emphasizes the key role of reflection and learning: movement from stage 1 to stage 2 must occur via stage 3. Whereas stages 1–3 are iterative and cyclic, stage 4 is isolated and visited only after the preceding stages are completed.

roles and influences of members of the research team in and on the design process and the resulting technology artifact. ADR structures the design process around cycles of building, intervention, and evaluation that mirror the cycles of planned intervention and reflection used in *action research*. Action research embraces disruption and action on the part of the designer as a means to learn about a problem [14], [16]. As a result, within the ADR methodology, the resulting artifact is considered an instantiation of the space, time, community, and process in which it is developed [29], [30], and is termed the *ensemble artifact* to reflect this quality.

Our own experience suggests that, whether consciously or not, many visualization designers learn through actions, using explicit methods such as technology probes [31] and data sketches [32], or more implicit approaches such as abstraction and visualization suggestions during interviews. Throughout the design process, visualization designers disrupt and influence both the target users and the problem context, while simultaneously being disrupted themselves. We suspect that adherence to the principles and process of ADR could increase the reliability of visualization design research by applying qualitative methods, such as those described in *grounded evaluation* [33], and providing guidance for capturing important details around insightful disruptions.

2.2.2 ADR Principles

ADR's seven guiding principles express the core values of the methodology and serve as a system of reminders to help ensure that research conducted by the ADR team, comprising visualization researchers and domain experts, is reliable throughout. Principles associated with early stage research (**P1**, **P2**) stress the importance of grounding design in theory and real-world problems, whereas those associated with design development (**P3**, **P4**, **P5**) focus on the influences of members of the collaboration on each other and on the design and on the need for continuous evaluation that takes such influences into account. Additional principles (**P6**, **P7**) emphasize the importance of acknowledging and responding to such influences throughout the design process, as well as the importance of generating usable design knowledge from specific research outcomes. Some of these principles (**P1**, **P2**, **P5**, **P7**) reflect those explicitly articulated in the visualization literature [7], [9], [10], [13], [34]–[36], but several of these principles (**P3**, **P4**, **P6**) provide new and potentially valuable guidance around the validation of visualization design research.

- **P1: Practice-Inspired Research.** The first principle emphasizes that applied research should be motivated and inspired by real-world problems. This notion is analogous to what the visualization community has termed problem-driven research. This approach helps to ensure domain relevance and paves the way for in vivo evaluation — two core tenets of design studies [13].

- **P2: Theory-Ingrained Artifact.** The second principle stresses the importance of design theory *and* domain theory in informing a design researcher's understanding of the problem and solution space and in helping guide the design process. This principle serves the same purpose as the *learn* stage in the design study methodology [13], which emphasizes that researchers must learn the space of visualization possibilities in order to design effectively. Additionally, it relates to the nested model's [10] *encoding and interaction threats*, which stress that theory should inform decisions at all levels of the design, thereby ensuring that resulting artifacts are theory-ingrained. Visualization design studies frequently focus on the application of theory to inform and justify design.

- **P3: Reciprocal Shaping.** Principle 3 emphasizes the constant shifting and shaping of both the artifact and the design process by the different perspectives within the team. This element of design research may feel familiar and perhaps obvious to members of the visualization community, but few attempts have been made to account for this element in existing visualization methodology [32]. Acknowledging the occurrence of reciprocal shaping can increase reliability by: 1) providing an explicit opportunity to document impactful activities and insights throughout the design process; and 2) revealing opportunities for structured approaches to ensure and support the effects of all people involved in the design process and resulting artifacts. We suspect that reciprocal shaping is more prevalent in applied visualization research than in information systems research due to the deeply collaborative and highly iterative nature of visualization design and the influential nature of data-led discovery. Recent work aiming to increase reliability by facilitating the documentation of design rationale in real-time throughout the design process provides a potential framework for ensuring and capturing reciprocal shaping [12]. However, further investigation and explicit guidance are needed.

- **P4: Mutually Influential Roles.** Principle 4 emphasizes the learning and cross-fertilization that occurs among ADR team members. Each member of the team brings a unique suite of knowledge, theory, and expertise. Through close collaboration, the team members learn about each other's expertise, sometimes offering valuable insight into another member's primary research domain. These insights can create substantial shifts in how another team member thinks about, or approaches, his or her research or domain, which within the visualization community is informally considered a sign of success for design studies. Beyond the citation of publications in an application domain, or anecdotal stories [1],[37],[38], however, methods and mechanisms for reliably assessing and reflecting on mutual influence are underdeveloped.

- **P5: Authentic and Concurrent Evaluation.** Principle 5 stresses that evaluation should happen throughout the design process to both influence the process itself and inform design decisions. Importantly, this principle encourages researchers to prioritize *authenticity* when evaluating artifacts and their effects through methods that are ecologically valid and conducted in-the-wild; this value is echoed in visualization design methodologies [7],[13],[36]. Authentic and reliable evaluation in the context of **P3** and **P4** — which value and encourage the influence of all members of the team on the shaping of the artifact and even in developing the insights achieved across domains — is, however, at odds with controlled studies that aim to establish predictive models through in vitro experiments that seek to remove subjectivity. Instead, ADR provides a means of achieving reliability within a subjective context by articulating the role of people and context in the design process itself. Furthermore, ADR, and visualization design research more generally, provide an environment for: 1) the evaluation of the *results* of controlled studies; and 2) a real-world scenario that may draw attention to the need for additional visualization research that may itself require controlled studies.

- **P6: Guided Emergence.** Principle 6 encourages researchers to be aware of and sensitive to the reciprocal shaping of theory-ingrained artifacts that happens throughout the design process: to nurture and incorporate the shaping into the design process, and to capture and apply it toward generating new design principles. Design research should be guided in part by theory, but researchers should also be open to incorporating insights that emerge from the research context, interactions, and evaluation. This principle helps ensure the reliability of the resulting artifacts by encouraging explicit awareness, and documentation, of the emergence itself, such as the evaluation of design decisions and the evidence used to develop them.

- **P7: Generalized Outcomes.** Principle 7, which recognizes the unique and highly specialized outcomes of a design process, also emphasizes the importance of generalizing and abstracting research findings. This principle specifically encourages researchers to generalize the problem and the solution, as well as to derive design principles. The visualization community subscribes to a similar process of generalizing research findings for the purpose of broader application. Abstractions, problem characterizations, and guidelines are examples of such generalizations [9], [10]. This principle, however, suggests that other kinds of learning, particularly surrounding the reciprocal shaping and mutually influential nature of the design process, could also be formalized, benefitting the greater research community.

2.2.3 ADR Stages

ADR is a high-level framework that encompasses many of the details found in existing visualization process models and practices. Research begins with preliminary investigation and articulation of the problem, continues with a period of iterative and cyclic human-centered design and development, and ends with critical reflection and synthesis of research. ADR differs from visualization design methodologies in its focus on intervention as a critical element of the design process and its objective of learning through design. Equally important, however, is the actionable framework that ADR's stages provide for adhering to and reflecting on the principles discussed in Section 2.2.2 to underpin the design process in ways that aim to achieve reliability. Tight cycles of action and evaluation are core to this and, unlike the emphasis in models of visualization design, reflection is required and ongoing.

- **S1: Problem Formulation.** The ADR process is triggered by a real-world domain problem, either expressed by domain experts or discovered by design researchers. The problem formulation stage involves the preliminary research and investigation of the problem, including narrowing in on the research opportunity. This stage also involves what ADR terms "casting the problem as an instance of a class of problems," similar to the initial problem characterization and abstraction of visualization design research. This stage emphasizes the principle of practice-inspired research (**P1**), stressing the importance of a real-world context for developing appropriate tasks as well as for establishing an ecologically valid context for validation of artifacts. The principle of theory-ingrained artifact (**P2**) is also stressed in this stage, indicating the importance of a prepared mind for developing effective solutions.

- **S2: Building, Intervention, Evaluation.** The second stage of ADR is grounded in the core tenet of action research that an effective way to learn about something is to try to change it [14]. In this stage, design researchers collaborate closely with domain practitioners, to develop and refine the problem space, and to design, develop, and refine the artifact, which is accomplished via cycles of building, intervention, and evaluation (BIE). As these cycles progress, new interventions are designed based on the results of previous cycles, are evaluated in real time, and are used to inform subsequent cycles. Technology probes [31] are a common intervention instrument used within design study research. Our own experiences conducting visualization design research suggest that BIE cycles occur frequently and at multiple scales, with overarching cycles exploring high-level questions, mid-level cycles exploring core concepts surrounding the data abstraction and design of a visualization artifact, and low-level rapid, iterative feedback and informal evaluation cycles throughout.

The principles of reciprocal shaping (**P3**) and mutually influential roles (**P4**) emphasize the highly collaborative, messy, human-centered nature of BIE cycles, as well as the shifting nature of the problem being studied. These principles provide structure to incorporate these dynamic and unpredictable elements of applied research into the design process. This stage also emphasizes authentic and concurrent evaluation (**P5**) as designers probe with technology to find out what works, and what their design ideas reveal. Evaluation needs to be quick, as well as concurrent with the build and intervene activities.

- **S3: Reflection and Learning.** The reflection and learning stage happens continuously and in parallel with **S1** and **S2**. In this stage, researchers are encouraged to reflect on ongoing evaluation in order to guide the design process; how well the research process adheres to guiding principles and how to encourage deeper adherence; and potential, broader implications of the research. This stage may occur either momentarily or in longer stretches, and is often triggered by an *insight*: a revelation, a moment of validation, or a design challenge developed during **S1** or **S2**. This stage has similar objectives to the *reflect* stage in the nine-stage framework for design studies [13], but ADR is explicit about the repeated and central role of reflection throughout the design process. The reflection and learning stage is guided by one principle, guided emergence (**P6**), encouraging researchers to adhere to **P2-5** throughout the design process and to reflect critically on the impact of such principles on the design and on the greater contribution of their research.

- **S4: Formalization of Learning.** The final stage of ADR is the formalization of learning. This stage occurs once the BIE cycles are completed and builds on the reflection and learning conducted throughout the design process, casting the insights and artifacts to a broader class of problems and solutions. Stage 4 embraces the generalization of outcomes (**P7**), pushing visualization researchers to think more broadly about the scope of their contributions to provide guidance around generalizing and abstracting elements of the design process. This stage serves a similar function as the *reflect* stage of the design study methodology.

2.3 Comparison to Visualization Models

As described in Section 2.2, ADR marks an evolution of thinking within the information systems community about the role of design in research and specifically about how to make design research reliable and generalizable. The visualization community is engaged in a similar evolution of thinking, and has put forth a number of high-level models for structuring the design process [7], [13], [39], [40] and validating design decisions [9], [10], [33]. In this section, we discuss these models as they relate to ADR.

Early visualization process models sought to define and analyze the existing visualization design space in order to guide the creation of new visualization systems [39], [40]. The desire to formalize visualization design was also reflected in a number of models and methods for evaluating visualization systems and approaches [36], [41]–[43]. Missing from this developing theoretical foundation was an explicit linking between the stages of the design process and appropriate methods of evaluation.

The nested model [10] was an attempt to bridge this gap. The model characterizes the visualization design process as four levels of research outcomes: problem characterization, data and task abstraction, visual encodings and interactions, and algorithms. For each level, guidance on the appropriate methods of evaluation is provided, and potential threats to validity are identified. In effect, the nested model guides researchers in their decision-making throughout the design process, emphasizing how faulty decision-making at one level causes threats to validity at all inner levels.

Extending the nested model, the nested blocks and guidelines model [9] provides mechanisms to capture and reason about context and decision-making at each level and structure for proposing generalized guidelines based on individual visualization designs. These extensions mark a shift in thinking toward increasing reliability and generalizability in highly contextual, specified, and dynamic research environments.

The nested model and its extension guide visualization researchers in decision-making, but they do not offer guidance on how the various stages of the design process should be carried out. The nine-stage framework represents the first formalized process model and methodology for conducting design studies [13]. Echoing the dual mission that drives information systems research, the methodology defines the aim of design study as generating new visualization knowledge through developing effective solutions to real-world problems. As with action design research, collaboration with domain experts is stated as a fundamental research component.

Drawing from the field of human computer interaction and from the social sciences, the nine-stage framework guides researchers through the messy, iterative, and highly collaborative design study process. The framework is organized into three phases: the first describing a set of activities that should occur before triggering a design study project, the second describing the core design activities in the production of visualization technology artifacts, and the third describing analysis and reflection to move design insights toward generalizable knowledge. Evaluation is stated as a concurrent step across the entire nine-stage framework, but its specific role or guidance on what types of evaluation are appropriate is not discussed in detail.

The design activity framework (DAF) [7] was a response to some of the shortcomings of the nine-stage framework. Specifically, the DAF emphasizes evaluation as a primary component of each design activity within the design process while also offering guidance for appropriate evaluation methods. The DAF also attempts to give a more flexible structure to the design process by supporting iterative, nested, and parallel design activities. In an effort to boost the actionability of the framework, the DAF bridges between the steps designers take and the decisions they make by explicating the levels of the nested model [9], [10] that are considered in each design activity.

ADR encompasses both the latter two stages of the nine-stage framework and the entire DAF by describing the design process from a problem trigger through formalization of the knowledge acquired. Similar to the DAF, ADR has an explicit treatment of evaluation as an essential step that is repeatedly visited throughout all design activities. Unlike the DAF, however, ADR makes reflection a primary activity throughout the design process, extending the role of reflection from that detailed in the nine-stage framework in a way that mirrors action research cycles. The benefits of concurrent reflection are gaining traction within the visualization community, and recent methods encourage researchers to critically reflect on process and decision-making [11], [12], but concurrent reflection has yet to be represented in high-level visualization models [44].

Additionally, the messy, iterative nature of applied design research is acknowledged by visualization decision and process models alike [7], [9], [10], [13], but very little guidance is available on how to engage in this fundamental element of research. ADR takes an important first step by emphasizing the importance of reciprocal shaping, mutually influential roles, and guided emergence.

The biggest shift that ADR presents over existing visualization models, however, is the adoption of action research [16], [45], impacting the design process in a number of ways: first, the emphasis on the role of learning through planned actions as a primary driver of the design process; second, the view that the development of an artifact is both a contributor to *and* consequence of the research process; and third, the framing of the design process in a manner that achieves reliability by incorporating established values from the social sciences.

CHAPTER 3

FORMATIVE DESIGN STUDY: POEMAGE: VISUALIZING THE SONIC TOPOLOGY OF A POEM

The first, formative piece of work in this dissertation is a design study with poets and poetry scholars. The initial goal of the design study was to develop visualization tools to support the close reading of a poem. We approached the collaboration as a standard design study, following the guidance of established design study methodology [8], [10]. Significant preliminary research was devoted to understanding and characterizing the domain problem, which turned out to be wicked [24] by nature; this was followed by a period of iterative prototyping and experimentation in order to better understand the data and tasks, and to explore potential visualization solutions. The results informed the design of Poemage, a visualization tool for interactively exploring the sonic topology of a poem.

By conventional standards, the design study was a success. Poemage proved to support the close reading of a poem. The tool facilitated important poetic insights and provided a new and engaging lens through which our collaborators could interpret and analyze poetry. Additionally, the tool implemented several novel visualization features and a novel data and tasks abstraction for the close reading of poetry.

These contributions, however, were not the most interesting results of the collaboration. Far more compelling from a visualization research perspective were the challenges that we encountered throughout the project, and the learning that stemmed from those challenges. When it came to publishing our results, we found that we had neither the methodological tools nor the confidence to formalize this learning in a way that was useful for the visualization research community. Instead, we chose the more traditional approach, putting forth the validated tool and its associated novel visualization components as the primary contribution, and including only a tentative discussion of our more interesting research findings. Reflecting on this experience caused us to question existing visualization methodology and to look to fields beyond visualization for potential guidance.

In this chapter, we present the details of the design study as they were originally presented to the visualization community. We then reflect on the project as a whole, on the major insights that we gained, and on the broader implications for applied visualization research.

3.1 Overview

The use of digital tools across disciplines in the humanities has exploded during the last decade. Popular projects such as the Google Ngram Viewer [46] and Wordle [47] have harnessed the power of computation to look across huge corpora of texts, leading to insights that had never been available before. Tools such as these are highly effective in supporting what is called *distant reading* — a term coined by literary scholar Franco Moretti to describe critical approaches that seek to understand literature and literary history by aggregating and quantitatively analyzing large text corpora [48].

Despite this new mode of scholarship, traditional humanities scholars continue to engage primarily in a very different type of analysis called *close reading*. As its name implies, close reading involves a detailed analysis of a text in all its complexity, encompassing an analysis not only of specific operations such as syntax, rhyme, and meter; such figures as metaphor and allusion; and such linguistic effects as affect, but also of how these operations interact across the temporal and spatial field of the text, with each other and with the reader, to create meanings greater than the sum of the parts. As this description suggests, much of the work done in close reading is well beyond the current capabilities of computation. Thus, the true value of computation to close reading is still very much in question and is the topic of an ongoing dialogue in the digital humanities. A handful of computational tools have been designed to support close reading, but much of the problem space remains unexplored.

We conducted a 2-year design study with poetry scholars and practitioners to explore this gap. Our two primary collaborators, both of whom are coauthors on the publication of this work, identify both as poets and as academics. We also engaged a network of practitioners, including two professors and two students of poetry. Together, these collaborators have literary expertise in medieval, early modern, modernist, and contemporary poetry, and they analyze poetry from a range of traditions and periods. Furthermore, they write formal verse, free verse, and experimental poems, and thus bring a diversity of theoretical viewpoints to their critical and creative work.

During this design study, we encountered several specific challenges that affected our design process. First, supporting close reading of poetry is a wicked problem [6], [24]: not only was it initially unclear as to *what* to visualize in a poem, but the design space for creating visual representations of poems and their features was completely open, since the use of technological tools as direct interventions in close reading (as opposed to in pedagogy and instruction) is still almost unknown to literary scholars. The second challenge we faced was that our collaborators belong to a community that sees the integration of technology into their research practices as largely unnecessary and potentially even intrusive. These challenges motivated us to use a highly collaborative and exploratory design process that takes the same experimental and even playful approach that our collaborators exhibit when reading and writing poetry.

Our design approach not only enabled us to learn more about poetry and close reading, but also disrupted our collaborators' view of poetry, pushing them to develop new perspectives on how poetic devices within a poem work together to create a response in the reader. These new perspectives led us to consider the *topology* of a poem, the complex structures formed from the interaction of sets of words across individual poems. Specifically, within a given poem we consider sets of words with similar sonic patterns. We focused our visualization design efforts on capturing poetic topology and providing a canvas for poets to explore the complex structure of sonic devices within a poem.

The specific contributions of this work are a characterization and abstraction for visualizing sonic devices in poetry; an open-source implementation of a tool for visualizing the sonic topology of a poem, called Poemage and shown in Fig. 3.1; validation of this design study through several case studies that illustrate the efficacy of Poemage for not only providing novel analysis insights but also enabling the creation of new poems and literary ideas; and a reflection on the unique nature of conducting visualization research in literary studies.

3.2 Previous Work

A number of highly effective tools exist in support of distant reading. Synoptic text visualization tools like GistIcons [49], Docuburst [50], Compus [51], and Galaxies [52] employ semantic analysis to extract key concepts and allow users to gain quick overviews of one or more documents and to run comparisons across large bodies of text. Tag cloud-based tools like Wordle [47], TextArc [53] and the variant Parallel Tag Clouds [54] provide a different kind of summary by focusing on the frequency and distribution of individual words or phrases. Several more sophisticated tools [55] [56] provide broad overviews while also allowing users to explore finer level connections. In general,

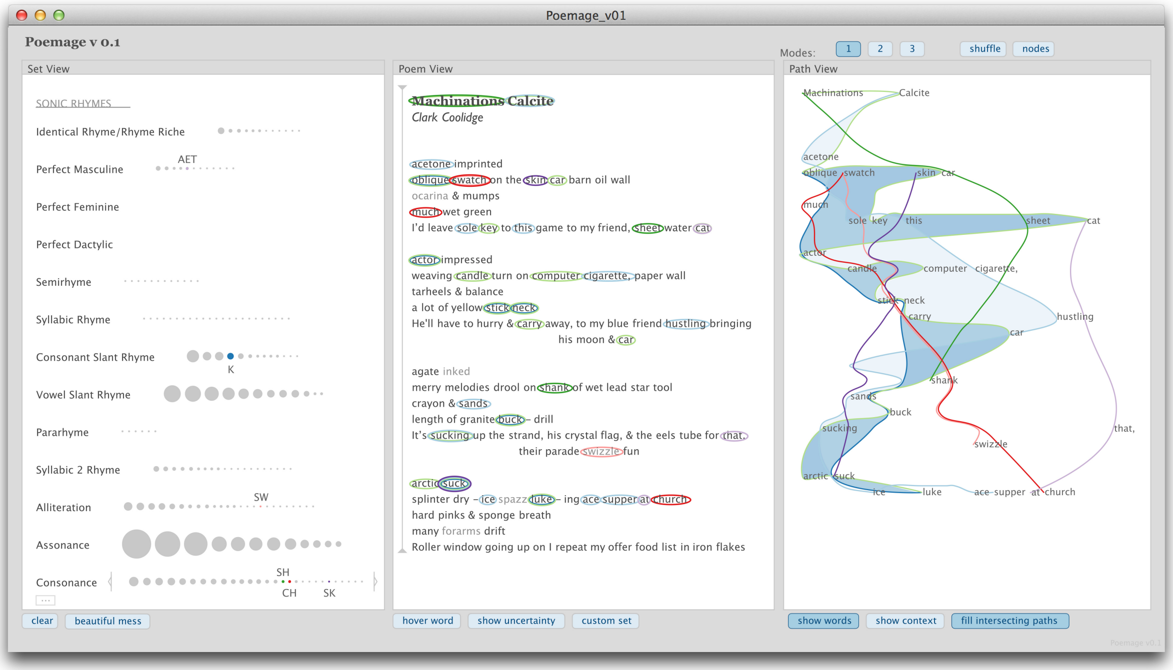

Fig. 3.1. The Poemage interface. The interface comprises three linked views: (Left) the set view allows users to browse detected rhyme sets (words linked through instances of sonic and linguistic devices). (Middle) the poem view allows users to explore the sonic topology directly via the text. (Right) the path view presents the sonic topology of a poem.

these tools treat a given text, or texts, as a bag of words, on which they perform a range of analysis without regard to structural and semantic context, features that are critical to the interpretation of poetic text.

Diverging from this slightly, FeatureLens [57] includes the repetition of expressions, revealing interesting patterns within and between documents. Techniques such as Phrase nets [58], Arc Diagrams [59], and The Word Tree [60] present more complex patterns based on a range of relationships. Building on The Word Tree, WordSeer [61] facilitates exploratory analysis of literary text.

Several visualization tools exist for specifically analyzing poetry. PoemViewer [62] employs rule-based visual mapping techniques to present a range of information about the poem, from traditional rhyme patterns to low-level sentiment analysis. PoemViewer attends to sound on a much deeper level than many tools, not only visualizing various types of phonetic repetition such as end rhyme, internal rhyme, assonance, consonance, and alliteration but also providing information about the physiology of sound production. PoemViewer provides a wealth of information, both structural and relational, but its interface does not capture the dynamism of a poem to the degree that our collaborators would like; poetic elements are, for the most part, presented as isolated objects, and poems are portrayed as static systems.

A second visualization tool, Myopia [63], was designed to aid in the close readings of poems. Myopia attends to a broad range of poetic elements, from meter, sound, and syntax to metaphor, personification, and emotion. These elements, however, must be coded a priori, a task that is currently done manually by a poetry expert and thus limits the tool's usefulness to a handful of poems. Although our analysis is limited to sound, Poemage processes text automatically, allowing users to explore any poem of their choosing, and for a broader range of sonic patterns.

The sonic analysis aspect of our research is closely related to Tanya Clement's seminal work on the analysis of aural patterns in text [64] and the exploration of the distance between the eye and the ear [65]. The visualization tool ProseVis [65] allows users to interactively explore the sonic transcription of a text and aids in the discovery of sonic patterns on different levels of granularity. Whereas ProsVis, and to a slightly lesser degree, Myopia and PoemViewer, capture and visualize the individual components of sound, Poemage extracts a range of more complex sonic patterns from a given poem and visualizes the interaction of such patterns across the space of the poem.

3.3 Design Process

Our primary collaborators had participated in previous visualization research [62], which acted as a first step toward overcoming their resistance to integrating technology into their own practices. Their resistance was rooted in part in an anxiety that the computer would inhibit the qualitative experience of the poetic encounter and in part a skepticism that it would be possible to visualize the interaction of any set of poetic features at a level of complexity that would allow them to make new and interesting observations. This initial project, however, left them deeply intrigued. Although there was some remaining resistance, they approached this design study with "skeptical enthusiasm" to see if it is possible to use technology to probe more deeply, beginning with a high-level investigation into sound, into questions of what makes a poem a poem and how a poem does what it does. Their larger goal is to create a tool that will be of use to the broader poetry community.

Integrating computation into the practice of close reading is a wicked problem [6],[24]. To quote a prominent critic of the digital humanities, Stanley Fish, "You don't know what you're looking for and why you're looking for it, how then do you proceed?" [66] Thus, we initially spent significant up-front time in joint conversations with our two primary poetry collaborators to determine what their goals were and how they imagined a visualization tool affecting their experience of a close reading. These sessions took place, at a minimum, on a monthly basis, were held at the University of Utah, typically lasted from 2 to 3 hours, and were often recorded for future reference. One primary collaborator, who was nonlocal for the majority of this collaboration, participated remotely via video conferencing.

The initial conversations were broad and open-ended: the poets did not have specific goals, they did not want a tool to "solve" a poem [67], and they described a wide array of poetic devices, such as affect, imagery, sound, pun, and metaphor, that they look at in a close reading. Our collaborators presented examples of interesting features and interactions within poems they had previously studied. In parallel, we investigated established methods for computationally detecting and analyzing the devices that most interested them. For many of these devices, the level of analysis that was of particular interest to our collaborators was beyond current technological capabilities, such as the detection of metaphor and imagery. The exception to this was sound, which is detectable, with some limitations, by established computational linguistics techniques.

Once we, as a group, decided to focus on sound, our attention turned to developing a system that would automatically sonify a poem. Building on existing approaches for the sonification of text, we developed a formalism for analyzing sonic devices in English-language poetry [17]. The formalism describes sonic patterns as rhyme, with the definition of rhyme being one that is both broad and flexible: rhyme is a poetic device that varies in definition from poet to poet. Our formalism includes a language for expressing a broad range of visual and sonic rhyme types and an associated ASCII notation designed for poets. In addition, we developed an open-source implementation of our formalism, a tool called RhymeDesign, initially as a platform to test and improve our formalism and eventually as a tool for poets to explore custom sonic devices in poetry. This software subsequently supplies the back-end to our visualization tool Poemage.

Even after determining what sonic data we wanted to explore, considerable design challenges and open questions remained. Close reading covers a broad range of tasks, encompasses varying styles of analysis, allows many different points of entry, and accepts an extensive range of sometimes radically divergent interpretations. In addition, our collaborators admitted resistance to integrating technology into their close reading. Thus, we also had to cultivate their trust, commitment, and enthusiasm.

A highly collaborative and exploratory design process proved to be critical in helping us navigate these challenges. We began by discussing the poets' experience with, and the results of, their previous visualization research. Next, we employed a number of different techniques in an attempt to clarify our point of entry. The first technique was an observation of a pair of close readings between our two primary collaborators, starting with the poem "Prayer" by Jorie Graham, followed by a close reading of "Night" by Louise Bogan. Close readings can be performed internally by one poet or externally as a conversation between two or more people. Throughout many of our future conversations, our collaborators returned to "Night" and other poems and picked up close readings in order to illustrate particular concepts, such as how sonic patterns can reinforce or undercut semantics. Other techniques for clarifying our entry point included studying an annotated poem from one of our collaborators, giving our collaborators a list of potentially interesting sonic devices that could be detected computationally and having them compile a list articulating the various sonic features that they were interested in exploring, and attending public poetry readings to better understand the nature and practices of the poets and poetry scholars.

Based on these activities, we ideated on a range of design possibilities to pursue, which we then developed into a set of technology probes [31]. We discuss details of these probes in Section 3.4. The probes were successful both in engaging our collaborators and in helping us better understand the problem space. We iteratively refined the probes over the course of several months based on extensive user feedback, both casual and via formal interviews, from our primary collaborators as well as our extended network of poets and poetry scholars. The incremental steps and the adjustments we made in response to their feedback and critiques helped the poets become familiar with the technology and also resulted in an interface that reflected their interests, aesthetics, and values. In addition, because our meetings were highly conversational and interactive, the poets generated poetic insights in our meetings on the fly, simply in response to developing and imagining the tool. This gave them confidence that the work, and eventually the visualization tool, would be useful to them.

Results from the technology probes formed our initial design ideas for the tool Poemage. These ideas were implemented into an initial prototype and presented to our primary collaborators. Based on casual feedback, we refined and improved existing features and added new features, the details of which are provided in Section 3.7.

3.4 Technology Probes

The technology probes were implemented in Processing [68] and combined into a single, multi-tabbed interface, shown in Figure 3.2. Users would load a poem of their choosing into the interface, which displayed the text of the poem, along with information about selected sets of sonic patterns. Following an initial development period in which versions of the probes were presented for informal feedback to our primary collaborators, the technology probes were deployed, along with written documentation, to four of our collaborators. The collaborators were given approximately one month to experiment with the probes, after which formal interviews were conducted. Interviews included brief observations of our collaborators using the tool, followed by questions surrounding approach, capabilities, and general usability. Interviews were recorded and transcribed, and one observation period was screen captured.

The initial goal of the technology probes was to explore the many different aspects of sound within a poem, as well as the role that sonic analysis plays in close reading. Using these probes, we experimented broadly in order to better understand, and to help our collaborators better understand, what kinds of sonic relationship they were interested in exploring in a poem. What we found led us to develop a broader understanding of rhyme, which we discuss further in Section 3.5. Furthermore, these initial probes indicated to us that our collaborators were not interested in exploring individual

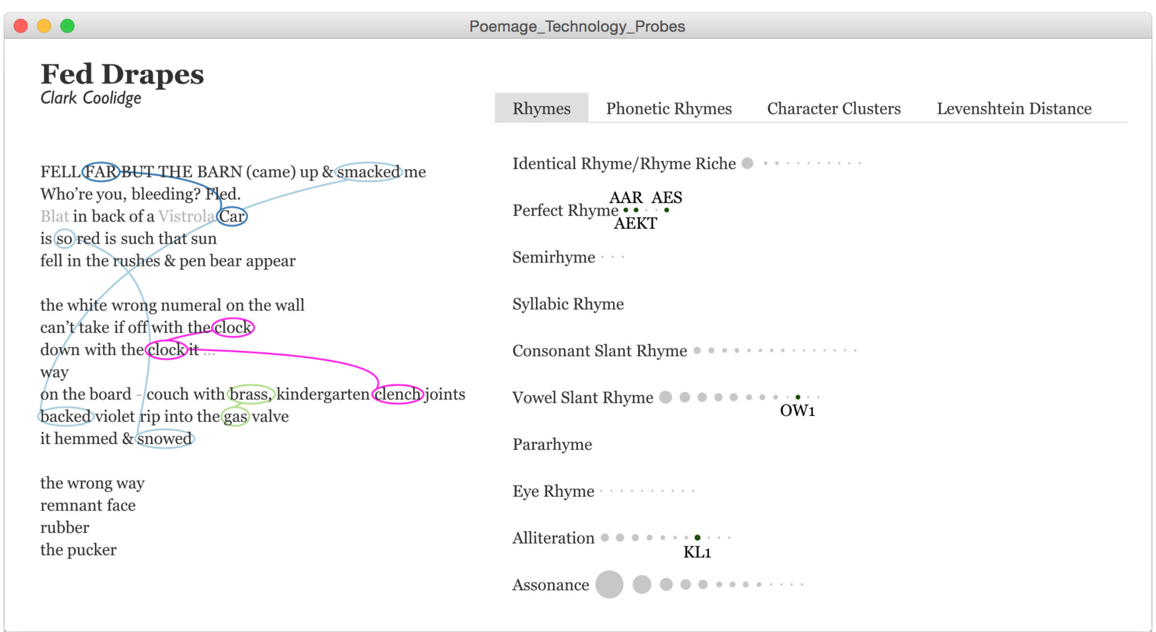

Fig. 3.2. Interface for the first set of technology probes.

sonic relationships, but instead they sought to understand how different sonic patterns interact and evolve across a poem. We thus developed a second set of technology probes to explore this notion of sonic topology. These investigations were instrumental to the development of our data abstraction, presented in Section 3.6.

The technology probes also allowed us to establish a common vocabulary with our collaborators; to focus on understanding how to capture data from a poem, as opposed to how to visualize it; and to define the space of what we could computationally detect in a poem. Overall, the probes helped us create an experimental and playful research environment that we maintained for the duration of the collaboration.

3.5 Poems and Sound

Poets and scholars see poems as living and relational, their literary features interacting not only with each other but also with us as readers. In close reading, a poetry scholar carefully attends directly to specific texts, tracing the interactions among such literary features as rhyme and meter, sound, figures, and syntax, while also considering how a given poem explicitly or implicitly converses with other poems in the literary canon. Although not viewed as an established technique for writing poetry, the experience of close reading often leads to the generation of new poems, and many poets do engage it as a prod to composition.

As a broad, literary device, sound provides poets with a rich source of play and can deeply influence the interpretation of the poem. Because of its emotional power and the way it works directly on the body of the reader of the poem, sound is an important source of poetic potency and can be used to reinforce or to undercut meaning conveyed via other poetic devices. In addition, sonic ambiguities — for example, in homographs such as *wind* and *bow* as well as in words with multiple pronunciations — also help generate multiple possible interpretations of the same poem. Furthermore, unlike many devices that may or may not be present at a particular moment or even at all in a given poem, sound is arguably pervasive in every poem at all levels.

Our collaborators consider a broad range of sonic and sound-related devices in their close readings of poems: from traditional types of rhyme such as *rhyme/sublime* and *picky/tricky*; to patterns involving the spellings of words, including anagrams (*desserts/stressed*) and eye rhymes (*cough/bough*), which may or may not relate sonically; to patterns surrounding the physiological production of speech sounds, such as the position of the tongue in relation to the palate. In this design study, we refer to all sonic and linguistic devices as *rhyme* [17], a broad definition embraced by our collaborators.

Our collaborators are particularly interested in the conceptual metaphor of a poem as a flow [69]. By approaching a poem, for the purposes of visualization, as a fluid moving via its linguistic devices and figures through a defined space, the flow metaphor captures three distinct levels of poetry analysis: the movement of individual linguistic and sonic devices through the space of the poem, how the interaction among such devices contributes to the complex sonic-temporal structure of the poem, and the impact that individual flows and collections of flows have on their surrounding region. In addition, flow introduces the notion of sonic turbulence — a metaphor for locations in a poem where there is increased intensity, energy, and activity due to the interaction of poetic devices. Over the course of our collaboration, we worked to translate this conceptual metaphor first into a data metaphor, expressed in terms of the extracted sonic patterns, and then into a visual metaphor, visually encoding the features and characteristics captured in our data metaphor. We discuss our data metaphor in Section 3.6 and our visual metaphor in Section 3.7.

3.6 Abstraction

In order to visualize the flow of a poem, we translate this metaphor into three data components: *poemspace*, *rhyme sets*, and *sonic topology*. The first component, **poemspace**, is the 2D space of the poem as it is printed on a page. Our collaborators were adamant about the importance of maintaining the spatial and textual context of the poem itself for two reasons. First, a poet can play with whitespace and layout to encode or enforce some sort of meaning in the poem. Second, any sort of data we pull out from a poem computationally will (usually) be meant to augment the reading of the poem itself. Poemspace, however, is unique compared to other 2D spaces as the reading of the poem constrains the way that movement within poemspace can happen: left to right, followed by top to bottom. In poemspace, each word has a location based on where it falls within a line, and where that line falls within a poem.

Words are related to each other not just spatially within poemspace, but also based upon their similarities to other words with respect to some sonic pattern. For example, in Fig. 3.1, the words *cat*, *that*, and *at* (underlined) are in a set together because they form a user-selected perfect rhyme. A set of words linked by a rhyming scheme is called a **rhyme set**. In this set, the words are ordered based upon their location in poemspace. Each word in a poem can belong to none, one, or many different rhyme sets, depending on which rhyme schemes are defined.

The **sonic topology** of a poem is represented by the distribution of rhyme sets across a poem and how those sets of words interact with each other, or not. From the conceptual metaphor of a poem as a flow, the places of sonic turbulence in a poem exist where multiple rhyme sets intersect, i.e., a word exists in multiple sets. To capture the sonic topology, we create paths from the rhyme sets, where each word in a rhyme set is connected by a link based upon its order in poemspace. For example, in Figure 3.1, the set including *machinations*, *calcite*, and *oblique* is ordered from top to bottom, left to right.

Our collaborators noted several different types of path interactions that are of interest, each of which is illustrated in Figure 3.3:

- *intersecting*: paths intersect at a single node.
- *overlapping*: paths intersect at multiple consecutive nodes.
- *merging*: paths intersect and then overlap.
- *diverging*: paths overlap and then split.
- *emerging*: paths begin at a point of intersection.

At a low level, our collaborators are interested in identifying and exploring places of turbulence, indicated as intersecting, merging, diverging, and emerging paths. At a higher level, they want to understand the places of turbulence within the context of the poem, and in the context of other poetic devices they identify in the course of their close reading.

As described in Section 3.5, sonic ambiguity exists for words with multiple pronunciations. The implementation of our formalism for describing rhyme captures this ambiguity and stores multiple versions of rhyme sets based on alternate pronunciations. For our collaborators, this ambiguity is

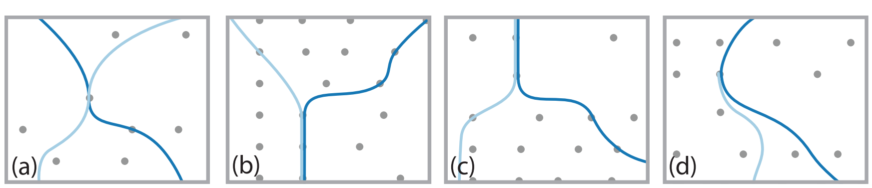

Fig. 3.3. Path interactions of rhyme sets. When rhyming sets are represented as paths, several interesting interactions can occur: *(a)* intersecting paths, *(b)* merging paths (also displaying overlap), *(c)* diverging paths (also displaying overlap), *(d)* and emerging paths.

a source of great joy as it represents possible alternate or additional meanings, and so enriches the interpretive experience of reading the poem. We can, then, describe the set of paths that results from sampling the ambiguous pronunciations as an ensemble of possible paths through a poem. Our collaborators are interested in exploring this ensemble to probe for different sonic topologies and different poetic interpretations.

3.7 Poemage

Our second contribution is the design and implementation of Poemage, a visualization tool for interactively exploring the sonic topology of a poem. The Poemage interface, shown in Figure 3.1, comprises three linked views: the set view *(left)*, the poem view *(middle)*, and the path view *(right)*. Multiple views provide users with multiple entrances into the poem, a feature expressed repeatedly by our collaborators in the technology probes as being highly effective for gaining new perspectives and insights. In addition, multiple views allow users to manipulate and play with the text, an example of which is presented in Section 3.8.2, and to view abstracted representations of the poem while maintaining a close connection with its original form.

A user's session with Poemage begins with the selection of a poem of interest, which is loaded into the tool via a text file. Poemage preprocesses the poem and creates the rhyme sets based on 24 rhyme types built into the tool. For the current version of the tool, we worked with our collaborators via the technology probes to define a set of rhyme types that captures the majority of interesting sonic patterns in a poem. Table 3.1 lists the types of rhymes currently supported within Poemage. We note that the back-end of Poemage includes a formalism for defining rhyme that can be modified through the open-source release of the software, thus supporting a definition of rhyme much broader than those available in other poetry visualization tools.

In this section, we describe the design and capabilities of the individual views and of the interface as a whole. As part of the path view description, we present two novel extensions of existing graph visualization techniques. Poemage was implemented in Processing [68], and the source code is freely available at **http://vdl.sci.utah.edu/publications/2015_infovis_poemage/**.

3.7.1 Set View

The set view allows users to browse through the various detected rhyme sets for a given poem. Each circle represents an individual rhyme set, the radius of which encodes the relative number of

Table 3.1. Rhyme types implemented in Poemage for creating sets of sonically related words.

Rhyme type	**Description**	**Example**
Identical rhyme	match in all sounds. Includes repeated words and homographs	pair/pair; pare/pair
Perfect rhyme:	matching stressed vowels sound and all proceeding sounds	
Perfect masculine rhyme	stress on the final syllable	rhyme/sublime
Perfect feminine rhyme	stress on the second to last syllable	picky/tricky
Perfect dactylic rhyme	stress on the third to last syllable	gravity/depravity
Semirhyme	perfect rhyme with additional syllable on one word	end/defending
Syllabic rhyme	perfect rhyme between stressed and unstressed syllables	wing/caring
Consonant slant rhyme	matching trailing consonants of stressed syllables	and/bent
Vowel slant rhyme	matching vowel sounds of stressed syllables	eyes/light
Pararhyme	matching leading and trailing consonants of stressed syllables	tell/tail/tall
Syllabic 2 rhyme	rhyme between initial stressed syllables	restless/westward
Alliteration	matching leading consonant sounds of stressed syllables	languid/lazy/line/along
Assonance	matching vowel sound (independent of stress)	blue/estuaries
Consonance	matching leading and/or trailing consonant sound (independent of stress)	shell/chiffon
Forced rhyme	perfect rhyme with imperfect match in final consonant sounds	shot/top/sock
Eye rhyme	spelling indicates perfect rhyme but sounds do not match	cough/bough
Character clusters	matching substring involving 1-4 characters	restless/westward
Mixed character clusters	mixed substring involving 2-4 characters	inlets/itself
Anagram	words formed out of the same set of characters	nights/things
Phonetic alliteration	leading consonants of stressed syllable match in mouth placement	pen/boy
Phonetic assonance	vowels of stressed syllables match in mouth placement	edible/anchor

words participating in a given rhyme set. Sets are organized by rhyme type and are ordered by decreasing set size to make the largest sets most visible. The rhyme types are organized into *sonic rhymes*, which involve matching patterns in sound, and *visual rhymes*, which involve matching patterns of alphabetic characters. Collapse and expand buttons located to the left of each category header allow users to omit either of these categories from their exploration. We observed the necessity to explore visual and sonic rhymes together, and separately, in the use of our technology probes.

The set view supports browsing of specific rhyming sets. When a user hovers over or selects an individual set or entire rhyme type, a pop-up label indicates the specific rhyme pattern for the set, and the associated rhyme sets are also displayed in the poem and path views. Color is used to link sets and rhyme types across views. For our color scheme, we rotate through an adapted version of giCentre's *12-class categorical paired* colormap [70]. The notion of browsing was something that our collaborators responded very well to in the technology probes. Not only did it come naturally to them, but it also provided a slightly abstracted way of exploring the poem while maintaining a close connection to its original form.

Clicking the *beautiful mess* button at the bottom of the set view selects all rhyme sets. This feature was first requested by our collaborators in one of the technology probes, and subsequently became one of their favorite features as well as a surprisingly valuable addition to the tool. Despite our initial hesitation to support showing all rhyme sets at once due to visual clutter, our collaborators were able to make interesting discoveries with this feature. One such discovery was the single, isolated pronoun *you* in the poem "This Is Just to Say" by William Carlos Williams. The beautiful mess revealed that *you* was the only sonically unconnected word in the poem, as shown in Figure 3.4. This insight was particularly powerful to our collaborators given the poem's occasion: to see *you*, the addressee and recipient of the ostensive poetic apology, excluded from the poem's many sonic relationships sharply heightened their sense of the poem's ambiguities. This example demonstrates how sound can work with and against semantics to elaborate readers' potential interpretations, and even affective experiences, of poems.

3.7.2 Poem View

We designed the poem view to support direct exploration of poetic devices in a poem's original form. Similar to hovering and selecting rhyme sets in the set view, users can hover over and select

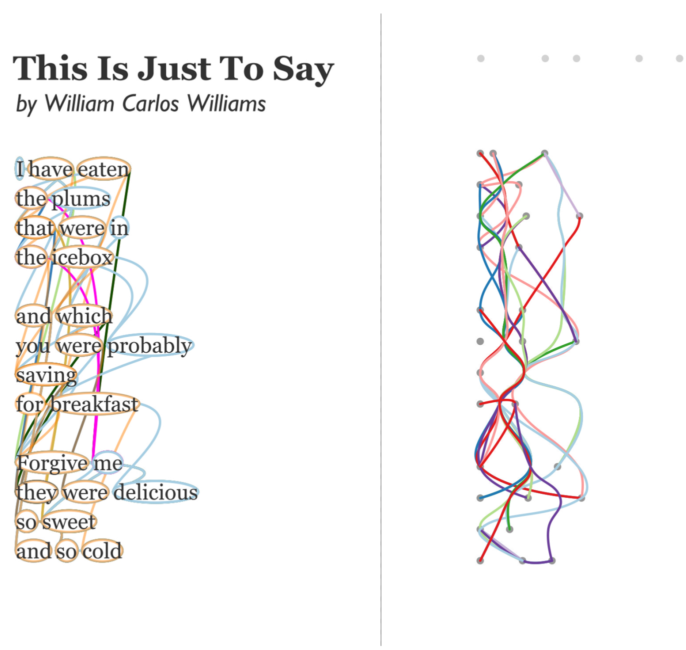

Fig. 3.4. The *beautiful mess* feature. In this example, the feature highlights the sonic isolation of the word "you" in this poem. A visualization from a technology probe obscures this isolation *(left)*, whereas rerouting in Poemage reveals the anomaly *(right)*.

words in the poem view, which in turn selects all the rhyme sets for which the word is a member. Browsing through words in the poem was requested by our collaborators early on and proved to be a very natural and effective way for them to interact with the text.

When rhyme sets are selected either via word selection in the poem view or directly in the set view, ellipses are drawn around the words in the selected set. The color of each ellipse corresponds to the assigned set color, as described in the previous section. Although we explored several different ways to encode selection, our collaborators preferred the ellipse, as it reminded them of their own annotation practices. For words belonging to multiple selected sets, concentric ellipses form a bullseye, similar to the concentric rings employed in LineSets [71], and provide a quick overview of the set membership for a given word. One collaborator commented that this encoding appeared to her as pebbles being dropped in a pond, with heavier pebbles causing more ripples, a nod toward a visual metaphor of the flow of a poem.

Clicking the *custom set* button allows users to build custom rhyme sets by selecting specific words in the poem view. Similarly, clicking the *show ambiguity* button highlights words with multiple pronunciations and allows users to select alternative pronunciations. Scrolling is enabled for poems of longer length with a print-to-pdf keyboard option providing a complete view of the poem and visualization.

3.7.3 Path View

The sonic topology of a poem is visualized in the path view, an abstract view that represents words in a poem as nodes at their corresponding location in poemspace. Our decision to map words to nodes, rather than to smaller linguistic units such as syllables, was rooted in observations of our collaborators during the technology probes and in the observed close readings consistently tracing sonic devices back to the semantics of the involved words. When a user selects a rhyme set in either the set or poem view, the associated path connecting the words in the rhyme set is shown in the path view as a curve connecting the associated nodes. We explored a variety of ways to represent the paths as node-link diagrams and found this representation best captured the characteristics of our flow metaphor.

We provide context in this abstracted view through several mechanisms. First, the path view is linked with the set and poem views such that selection and highlighting in the other views causes paths to appear in this view. Second, when a user hovers over a node in the path view, the corresponding word appears as a pop-up. Third, clicking the *show words* button displays the set words associated with a given path. Fourth, clicking the *show context button* displays nearby words surrounding a given path. The extent of surrounding words can be adjusted via the *context slider*. An example visualization employing these features is shown in Figure 3.5 *(c)*.

Rendering paths in poemspace requires a number of design considerations: rerouting paths to avoid ambiguous set membership of words, effectively rendering multiple paths at once, supporting multiple interpretations of poemspace, and incorporating sonic ambiguity. We discuss each of these considerations in more detail below.

We developed a number of prototypes to explore design variations for this view, starting with images we shared with our collaborators generated from off-the-shelf node-link diagram tools [72] [73]. Based on feedback about these tools, we designed a path visualization that resembles other line-based overlay techniques, such as LineSets [71], Kelp diagrams [74], and KelpFusion [75], in which the spatial context of the set data is preserved, and shortest path algorithms are employed to determine routes linking set members. Like KelpFusion, our approach combines line-based with region-based overlay techniques [76] [77] [78]. Rather than use convex hulls to delineate sets, however, our path visualization employs fill to emphasize regions enclosed by sets intersecting at multiple nodes.

3.7.3.1 Rerouting Paths

In our prototypes, we encountered problems with edges intersecting words that were not included in the path. Furthermore, examples like the *you* anomaly that we discussed in Section 3.7.1 highlighted the importance of having isolated nodes appear isolated. We therefore decided to *reroute* edges such that they explicitly avoid words not included in a path.

We experimented with several different rerouting [79] and bundling [80] [81] techniques before finalizing our design. In the process, we discovered that we could take advantage of the vertical, regular spacing of poemspace to establish a simple and general rerouting technique. Our technique reroutes edges connecting words that are separated by more than one line in the poem, as these are the edges that may intersect words *not* in the path's set, as shown in Figure 3.5 *(a)*. For these edges, at each line of the poem that the edge intersects, we determine the closest *whitespace* to the edge intersection, i.e., the closest space between words. We place a new control point at the center of these whitespaces and render the edge as an interpolating cubic Bézier curve. This rerouting produces a meandering curve that avoids all words that are not included in the path, illustrated in Figure 3.5 *(b)* and *(c)*. A side-effect of this rerouting technique is that similar paths are naturally bundled together.

This rerouting technique integrates several different aspects of the conceptual metaphor of a poem as a flow: the notion that adjacent flows tend to aggregate, intensifying the same path, as well as the idea that flows can behave like eddies, bending and diverging dominant courses, disrupting their surroundings, looping backwards, dissipating, or developing in new directions. In addition, the rerouting generates much more organic, aesthetic curves than those generated in our previous implementations, which increased the overall efficacy of the tool for our collaborators. Although we appreciate that minimizing wiggles is a common constraint in graph drawing, we claim that our approach improves, rather than obscures, our visual representation of poemspace. Our rerouting technique is most similar to techniques that use grid-based rerouting to bundle edges and reveal high-level patterns [82] . The inherent grid-like qualities of our data make our rerouting technique a more intuitive approach and allow us to avoid issues associated with standard grid-based methods.

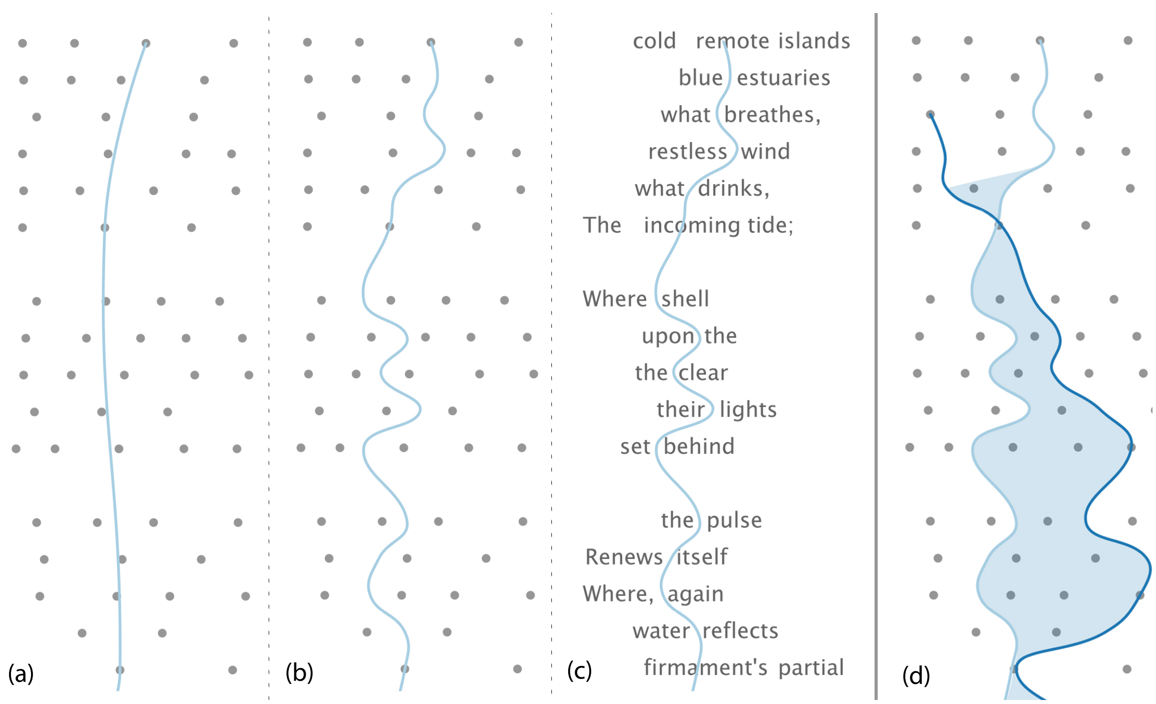

Fig. 3.5. Rerouting paths in poemspace. Rendering the rhyme set paths in poemspace can lead to ambiguous set membership *(a)*. Poemage incorporates a path rerouting technique to disambiguate words *(b)*, as well as context information for the nodes *(c)*. *(d)* Fill function between two intersecting paths.

3.7.3.2 Drawing Multiple Paths

As we discussed in Section 3.6, of particular interest to our collaborators was exploring places in the poem where paths overlap, merge, diverge, and emerge. Our path rendering algorithm supports visualizing these locations through two mechanisms: maintaining a consistent ordering of paths to help users trace paths across a poem and a fill-function that emphasizes mergence, divergence, and emergence.

The ordering task closely resembles the metro-line minimization problem [83] with a unique combination of constraints: rerouting, Bézier splines, and intermediate nodes. We adapt an existing technique to address these differences [84], first calculating ordering for pairs of paths sharing common subpaths and then iteratively adding one path at a time such that the pairwise orders are maintained. Path order is computed on-the-fly as users select and deselect paths of interest. We disable ordering for the *beautiful mess*, as it causes significant visual clutter and obscures isolated nodes left exposed via rerouting.

In addition to interactive path ordering, we include a semitransparent fill function to emphasize path mergence, divergence, and emergence, as shown in Figure 3.5 *(d)*. In places where paths intersect at multiple nodes, the fill spans the area enclosed by the two paths, and in places where two paths merge, diverge, or at single points of intersection, the fill approaches the intersection point. This fill technique is inspired by our collaborators' belief that interacting flows influence their surrounding region (and vice versa), sometimes pulling surrounding words closer together, other times pushing them apart. The fill seeks to reveal possible regions of influence formed via the intersection of paths. Fill is computed interactively on a pairwise basis. The user has the option of setting the fill to a constant color or to a color that is dependent on the involved paths.

3.7.3.3 Deforming Poemspace

Another design concept that we experimented with was that of added whitespace. Inspired by poets like e e Cummings and Charles Bernstein who use whitespace to augment the shape of their poetry, our collaborators found that comparing the shapes of paths with and without added whitespace helped form interpretations as to why an author may have formatted the text in the way he or she did. In the path view, we support three deformations of poemspace: the original form, compressed whitespace to just a single character width, and evenly spaced nodes. We illustrate these deformations in Figure 3.6. The path view supports toggling between deformations.

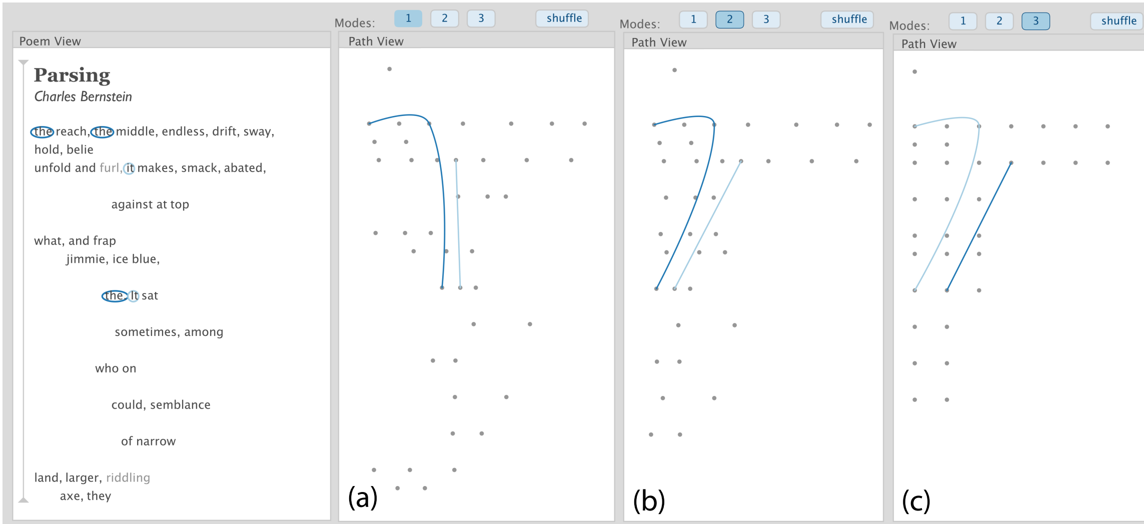

Fig. 3.6. Different modes of poemspace. The three modes of poemspace shown in the context of the poem "Parsing" by Charles Bernstein: *(a)* original form, *(b)* compressed whitespace, *(c)* and even spacing.

3.7.3.4 Ambiguity

The final feature in the path view addresses the concept of ambiguity. In addition to allowing users to select alternative pronunciations for homographs and other multipronunciation words in the poem view, a *shuffle* button reruns the entire program based on randomly selected pronunciations, thereby sampling the ensemble members described in Section 3.6. This shuffle is meant to visualize and inspire different interpretations of the same poem.

3.8 Validation

We present two forms of validation for this design study. First, four case studies with our collaborators illustrate how Poemage not only supports novel analysis insights (Section 3.8.1), but also how the tool supports making and remaking new poems (Sections 3.8.2 and 3.8.3). Poemage was introduced to the poets via demos highlighting the various features and interactions, many of which were previously familiar to them from earlier prototypes. These demos were either given in person or recorded. Following the demos, the collaborators were given a week to experiment with the tool, after which interviews were conducted, recorded, and transcribed to gather user feedback. Although semistructured interviews were planned, in all three cases the opening sequence of questions "Can you walk me through how you used the tool?" and "Did you gather any new insights, and if so, can you show me how you arrived at them?" propelled a dialogue in which the remaining questions were answered and many additional topics were approached. Two of the poets kept journals of their experimentation [36], which provided direct narration for three of these case studies.

As a second form of validation, we discuss the impact that our collaboration has had on the scholarship of our direct collaborators (Section 3.8.4); we argue that disrupting the thinking of these poets is an important mark of success for this work.

3.8.1 Close Reading With Poemage

One collaborator described her approach to using Poemage in analyzing a poem as "noodling," hovering over one sonic feature after another in the set view and poem view, selecting and deselecting rhyme sets almost arbitrarily. She said her greatest successes and insights came in every case when she happened on something indirectly, through idle play — as she says, "almost out of the corner of my eye."

A specific example of this was an insight gained when glancing at the placement of nodes in the path view for the poem "Night" by Louise Bogan. While the placement of nodes in this poem is mostly regular in that there are generally a similar number per line (around 4) and they are mostly at similar distances from each other, indicating that each line typically has the same number of words and these words are of similar length, one line had only two nodes, the second following very closely on the first. Thus, the abstracted view of poemspace revealed an immediately visible anomaly in the spatial distribution of words. This anomaly coincided with a powerful semantic moment in the poem, leading this collaborator to explore the rich sonic turbulence at that location and its connection and reinforcement of the semantic flow of the poem. She said that this view shed new light on a poem with which she was deeply familiar:

> In other words, not only is this the poem's turn, its pivot and crisis, but there's just a whole lot going on, a lot I wouldn't necessarily have considered in quite this way without the tool drawing my idle eye — a lot I hadn't in all these years considered up to this moment.

She commented that Poemage took her into the poem in a different way than she was accustomed to, and that this occurred via both the poem view and the path view. In a typical close reading, she begins with the title and first line and forges a slow, recursive path in which the overall movement is left to right and down the page, but in which a specific poetic event might send her back to the beginning of a line or up the page again as far as the title. An example of how Poemage changed this procedure for her occurs in the previously described observation of the node anomaly, which occurs late in the poem. Because of that observation, this specific encounter with the poem began with the first line of the final stanza, rather than at the beginning.

Another collaborator chose to explore Jorie Graham's poem "Reading Plato." She began by hovering through words in the poem view to see what overall patterns appeared in the path view. By scrolling up and down through the length of the poem, she was able to piece together a composite sense of the sizes and shapes of these patterns as they appeared and developed in the path view. Because she felt least likely on her own to discern specific examples of complex visual rhymes, she next turned her attention to that category of rhyme sets in the set view. She browsed through the mixed four-character cluster rhyme sets, which immediately revealed some interesting results potentially relevant to her interpretation of the poem. A visualization of her selections is shown in Figure 3.7(c). She reflected on her exploration:

> The multiple-view interface felt engaging and responsive and it reflects the sensibility that I experience when reading a poem: that interpretive readings are made, choice by responsive choice, and that nothing is absolutely conclusive. Curves and complementary soft colors, mixed and blended through interaction, connote changeability and invite engagement without visually overpowering the user. The poem itself remains central both figuratively and literally while the multiple, flexible paths through the poem allow users to shift their focus quickly, between minute details and single patterns in isolation, or in relation to each other and the poem as a whole. In these ways, Poemage not only reveals patterns within the poem, but also enables users to see their own spontaneous choices, their own interpretive work and critical explorations in new ways — which in turn spur still further exploration.

3.8.2 Erasure Poetry

Erasure is a form of poetry generated by erasing words from an existing text, resulting in a new poem with potentially new meaning. The concept of an erasure was introduced by a collaborator from our extended network of poets in response to one of our technology probes. Iterating on the idea subsequently led to our inclusion of the *show words* and *show context* features in Poemage. These features allow users to explore all the possible erasures formed from single or combined rhyme sets and their surrounding regions in poemspace. This collaborator experimented quite a bit with the final implementation of these features in Poemage, and recently exhibited several of her erasure poems generated using the tool in a local art gallery, one of which is shown in in Figure 3.7(a). We describe her experience using Poemage to generate erasure poems here.

This collaborator took several different approaches to using Poemage. For poems she was deeply

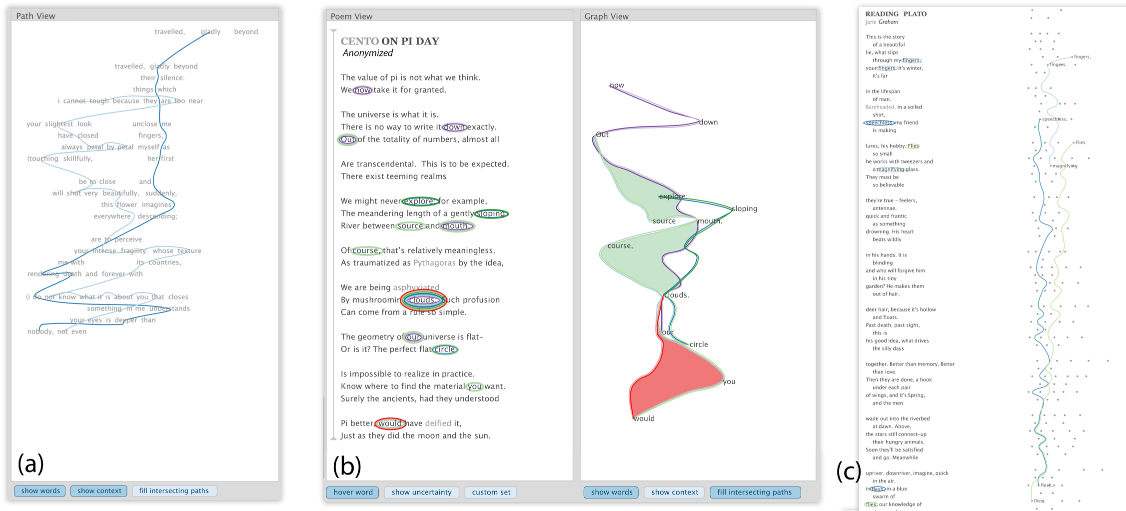

Fig. 3.7. User visualizations. (a) An example of an *erasure* of e e Cummings poem, “somewhere I have never travelled, gladly beyond” created by one of our collaborators and exhibited in a local art gallery. (b) A visualization showing a *cento* created by one of our collaborators. (c) A visualization of the poem “Reading Plato” by Jorie Graham, printed to pdf.

familiar with, her exploration was guided by previous observations and investigations, leading her to hover over and select particular rhyme sets in the set view. For less familiar poems, on the other hand, hovering over different words in the poem view, thereby revealing a word’s various sonic connections with other words in the poem, helped her to quickly gain entrance to the text. Taking a slightly different approach, she also experimented with building interesting shapes in the path view by randomly selecting words and rhyme sets in the poem view and text view. Such selections were based on their visual impact on the visualization in the path view. This third approach allowed her to specifically investigate how introducing new sonic flows changed the sonic structure of the poem. In a similar fashion, she generated new erasure poems by enabling the *show words* feature and selecting one or multiple rhyme sets based both on the shape of their paths and the subpoems they revealed. She commented that the visualizations, and especially those using the fill function, went “straight to the pleasure center of [her] brain.” She also commented that Poemage encouraged her to spend more time with the poem, which she felt was one of the biggest benefits of the tool.

3.8.3 A Cento in the (Re)making

A *cento* is a poem composed entirely of lines or passages taken from other authors. One collaborator used Poemage to explore a cento that she had composed by taking lines from an article in the *New York Times* written in honor of Pi day [85]. Her cento is shown in Figure 3.7(b). She loaded the cento into Poemage and proceeded, according to her usual practice, to noodle around, looking for various densities or lack of density. She also compared the cento to other poems entirely of her own making. What Poemage helped reveal to her was the extent to which she had managed to make another's text her own — how the poem "looked" sonically like one of hers. She recounts this:

> I noticed that my cento is in some ways as sonically intense as my poems built from scratch — with the notable difference that the sentences of a journalist tend to rely heavily on the verb "to be," a verb I use quite rarely when left to my own devices. "Is" shows up in the cento as the dominant sound on exact rhyme, assonance, and consonance; never in my own drafts. This initially made me a little despondent — should I give up making centos from the New York Times? — but another view of the poem shows me a wonderful set that included "now," "down," "out," "source," "mouth," "course," "clouds," "you," and "would," and even encompassed "circle," which feels just like me. This raises a question: how is my own cento (I have several of them) like me and not like me? How do I unconsciously select sentences for the cento not only for their meaning but also for their sound? How does a cento in the making become more like me as I make it?

This revelation, combined with other observations made using Poemage, inspired this collaborator to "sonically reload" her cento, rearranging the lines and passages based on the resulting visual representation in Poemage's path view.

3.8.4 A Disruptive Technology

One interesting, yet difficult to capture, measure of success for a design study is the act of disrupting the thinking of a domain expert [13]. Through the course of this design study, our poetry collaborators developed new perspectives on their domain and research practices through the lens of computation and visualization. Asking them to define precisely what is interesting in a poem, in a rhyme, and in a sound led to new thinking, which in turn enabled them to envision new ways of approaching a poem and to narrow the scope of their tasks from "find everything interesting in a poem" to "visualize the sonic topology."

One such insight was the rearticulation and development of the conceptual metaphor of a poem as a flow, into data and visual metaphors. As part of a session entitled "Things My Computer Taught Me About Poems" at the 2014 Modern Language Association's annual conference, our two primary collaborators presented on this metaphor, and specifically on the notion of turbulence and poetic time [86]. These collaborators also published significant articles in which they reflect on the impact that visualization research has had on their poetry scholarship. In one of these articles, the author focuses not on the technology itself but on how the need to teach the technology, and the computer scientists, what to look for and visualize in poems forced her to be much more precise in her own thinking about not only sonic devices and how they signify within poems, but also more complex questions of language and imagery [87]. The author of another of these articles describes how this collaborative research prompted her to re-read familiar poems in new ways long before our software was ready to explore, and discusses ways this research has led her to re-evaluate and reimagine her theoretical positions on such literary questions as how poetic time operates and even what reading entails [69].

We put forth these results as an important validation of the impact this work has had on our collaborators' poetry scholarship. Since these interviews, we have continued to work with these poets to explore new extensions of Poemage that probe more deeply into the literary concepts discussed in this chapter, as well as into new concepts such as sonic depth and the role of technology in promoting creativity.

3.9 Discussion

Working with poetry scholars has been a delightful, and challenging, process. Here we reflect on some of the issues that we believe make literary studies different from other, more traditional visualization problem domains. We offer a number of insights we gained as visualization designers and provide several suggestions for future work.

3.9.1 Breaking Convention

This research challenged us to embrace concepts that visualization convention seeks to resolve, specifically ambiguity and visual clutter. In the field of visualization, avoiding ambiguity is the norm. In this research, however, our collaborators led us to regard ambiguity as a fundamental source of insight. While Poemage includes some features that allow users to explore ambiguity within the data, we plan to investigate this topic in greater detail in our future work.

This work also challenged us to embrace a degree of visual clutter. The *beautiful mess* upends established visualization principles that value clarity and readability. This messy view, however, was consistently one of our collaborators' favorites, and it revealed one of the more important poetic insights of this work, shown in Figure 3.4. Our collaborators told us they would not have made this insight without the beautiful mess. In an interview, one of our collaborators commented that the beautiful mess was completely representative of what she and other poets seek to understand in a poem, namely how the constituent parts of a poem work together to form complex meaning. This collaborator also reported that in times of feeling overwhelmed by the technology, she turned to the beautiful mess to ground and re-energize herself. We wonder, however, if there is a degree of novelty in the beautiful mess that may wear off in time — we plan to revisit the utility of this view in the future.

In general, coming to terms with how quickly the visualizations became cluttered, and *not* restricting the tool to avoid such clutter, was a challenge for us as visualization designers. To our collaborators, this clutter, which they identified more as *chaos*, was inviting and energizing and was a space that they felt very comfortable exploring. Similarly, our collaborators' excitement about ambiguity as an aspect of the data that enhances meaning, rather than clouds it, caused us to reconsider how to include it in Poemage.

This experience taught us to be willing to put some of our own design principles aside and to be open to experimenting with unconventional visualization if explicitly, or implicitly, requested. Doing so led us to include features that our collaborators were genuinely excited about and also helped us to better understand the problem space. Throughout this research, our collaborators actively challenged and probed their resistance to integrating technology into their practices in the hopes of advancing their research, and we as visualization designers learned to do the same. This is a lesson that we plan to bring to future collaborations, and we encourage others to do the same.

3.9.2 A Screwmeneutic Approach

Within the digital humanities, an adventurous wave of research rejects the notion of using computation to solve a text or to verify existing hypotheses, and instead focuses on using computers to further literary criticism, to generate an indefinite number of unique and sometimes radical interpretations and to guarantee continued meaning making. Concepts of text *deformation* [88] and *tamperings* [89] have energized members of the literary criticism community and have motivated a somewhat informal branch of text interpretation delightfully termed *screwmeneutics*. This term comes from an influential paper by Stephen Ramsay entitled "The Hermeneutics of Screwing Around..." [90]. Tools created with a screwmeneutics sensibility encourage a certain amount of playfulness and creativity from their users. We embraced this philosophy whole-heartedly throughout this design study, from our approach to conducting research to the design of technology probes and Poemage. Our validation results in Section 3.8 explore a range of possible outcomes on the part of our collaborators that can emerge when poets are given a tool that supports free-form exploration, browsing, and play.

Because a significant component of this research was investigating the role and impact of technology on the experience of close reading, and because our collaborators gained so much from this aspect of the design study process, we wanted to allow users to conduct similar investigations using our tool. In addition to being a culmination of our research findings, we see the various components of Poemage additionally as technology probes for end-users — opportunities to explore the impact of technology on their individual reading of a poem.

3.9.3 Measuring Success

An underlying challenge throughout this design study has been determining how to measure our success. We attribute this to several different complications. First and foremost, the range of valid interpretations makes comparing against ground truths fairly unproductive. In terms of proper evaluation, our research fits in the *evaluating visual data analysis and reasoning* scenario [35], which encourages insight-based evaluations. In our case, in which gathering insights is a fundamental component of our collaborators' research, this method seems less indicative of success than it perhaps might be for a different domain. Therefore, in addition to highlighting specific insights gained using our tool, we also evaluate new *kinds* of insights and how the insight gathering process may have changed using Poemage, which connects back to our investigation of the role and disruptive impact of technology on close reading.

In reflecting on the outcome of this research, the question arises: could randomly selected sets of words yield equally valuable and equally abundant insights? Our collaborators have hinted that the answer to this question might be *yes*, which leads to a second question: what is the role of technology in a creative pursuit like poetry? We are continuing to explore these ideas with our collaborators and are designing user studies to test our hypotheses.

In a similar vein, we acknowledge that the findings and contributions presented in this chapter place a stronger emphasis on the role and impact of technology on poetry scholarship than on the support of close reading with Poemage. Although we are continuing to improve Poemage, we believe that developing a deep understanding of the intersection of technology and the humanities is fundamental to creating truly effective tools supporting a broad range of literary studies practices, thereby extending beyond close reading.

Finally, in this research, *pleasure* and *enjoyment* are productive research outcomes and play an important role in guiding exploration [69] [87] [90]. One consequence of this, and as we found through our validation, creating a visually pleasurable research environment goes beyond general aesthetics, encouraging richer exploration and greatly increasing the overall efficacy of the tool. In reflection, we call for an adapted set of evaluation guidelines for conducting research in the humanities, and specifically in literary studies. We believe that such guidelines could extend to other arts as well.

3.10 Reflection

Poemage proved effective in supporting the close reading of poetry; however, results of the design study extended far beyond the tool and its associated visualization contributions. We believe that far more valuable and transferable were the broader insights that we, the visualization designers, gained about conducting applied visualization research, the first of which surrounds the importance of *really* listening to your domain collaborators. Domain experts come with their own backgrounds, perspectives, and relationships with data, visualization, and technology. For example, listening to our collaborators' pleas for a cluttered view engaged them on a new level and led to one of the more important poetic insights of the collaboration. We anticipate that, in general, understanding and attending to these contextual factors, including expert hesitations, inclinations, value systems, and factors for engagement, can enhance the design process, and can also reveal new paths of inquiry and new visualization research opportunities.

The second major insight surrounds the role of disruption in applied visualization research. Borrowing a page from action research methods [15], we found that disrupting the way domain experts think about and interact with their data provides a powerful tool for discovery and insight. Visualization researchers can similarly benefit from disruption to their own research process and thinking, and design study collaborations provide a rich and natural environment for disruption and cross-fertilization. Along similar lines, we see great benefit in embracing friction within visualization design research and in seeing challenges as research opportunities rather than as road blocks.

The third major insight surrounds the notion of visualization research systems as dynamic technology probes [31], rather than as static implementations of visualization solutions. Such probes offer valuable platforms to experiment, investigate, disrupt, and to learn with. Additionally, we found that taking a more flexible and playful approach to system design can greatly assist in establishing trust and increasing engagement from domain experts.

These insights are examples of the broader kinds of learning that can stem from design study and that have the potential to advance applied visualization research. Such examples are indicative of the fact that much more learning occurs in design study than what we are currently capturing, leveraging, and communicating back to the visualization research community. In this design study, the learning surrounded the highly collaborative, contextual, and dynamic nature of our research. How to properly conduct this learning, how to validate it, and how to transform it into transferable visualization knowledge are open questions. Missing from existing applied visualization methodology is a framework that incorporates this fundamental dimension of applied visualization research.

CHAPTER 4

APPLYING ADR TO POEMAGE

Our experience conducting the Poemage design study caused us to question existing applied visualization methodology and to look to fields beyond visualization for guidance on conducting design research in highly collaborative, contextual, and dynamic environments. In this dissertation, we look to information systems research for insight. In particular, we turn to *action design research* (ADR) [5], which through adherence to a set of guiding principles offers a framework for reliably structuring and reporting on the design process in ways that can contribute to the acquisition of knowledge. ADR shares many commonalities with existing visualization design methodologies, but deepens the theoretical underpinnings through its use of *action research* [15] as a basis for design research. This foundation — adapted from an established method of inquiry in social science, in which researchers directly influence the context they study through planned intervention — affords a new perspective on the forces that shape the nature of visualization design, and on the way we define reliability of research.

In Chapter 2, we describe the ADR principles and stages using visualization parlance, and compare ADR to existing visualization design models. In this chapter, we present and critique the Poemage design study by reframing it through the lens of ADR. We describe the Poemage design process according to the four stages of ADR: *Problem Formulation*; *Building, Intervention, and Evaluation*; *Reflection and Learning*; and *Formalization of Learning*. For each stage, we reflect on our adherence to ADR's associated guiding principles.

In doing so, we find that ADR provides a valuable framework for interpreting and implementing the major insights from the Poemage design study, which surrounds the role of research context, disruption, and technological intervention, and provides initial guidance on structuring and formalizing a wider span of learning. We argue, more broadly, that this reframing sheds new light on contributions of the design study and illustrates the applicability of ADR to visualization design research.

4.1 ADR Phase 1: Problem Formulation

The Poemage design study was triggered by the poetry scholars' interest in exploring the potential role of visualization in poetry scholarship and, in particular, the experience of a close reading of a poem. Close reading involves the in-depth analysis of a poem and all its literary devices, which is central to the poetry scholars' research. During the initial portion of the problem formulation stage, the visualization researchers conducted informal and semistructured interviews with the poetry scholars to learn about the close reading process [33]. From these interviews, the visualization researchers discovered that influencing close reading could happen in many different ways, and that there was no explicit notion of data in this context.

From there, the visualization researchers dug through the text analysis literature to determine what types of literary devices — metaphor, imagery, affect, and sound, to name a few — can be extracted from a poem. The goal was to find a device that was both robustly computable *and* interesting to the poetry scholars. Eventually, the team narrowed the choice to sonic devices, a class of poetic device that utilizes sound and the relationships between sounds in words to effect the interpretation of a poem. Next, the team developed an initial data abstraction; the data under consideration would be sets of sonically similar words within a poem. Additionally, the visualization researchers reviewed literature around text visualization, close reading, and digital humanities.

4.1.1 Adherence to P1 and P2

The investigation in this stage was grounded in the poetry scholars' interest in exploring the role of visualization in poetry scholarship, which was continually revisited through the numerous discussions among the team members. The theory acquired within this stage came from various approaches to text analysis and visualization, and the values and methodologies of digital humanities. In addition, learning about approaches to digital humanities scholarship inspired the visualization researchers to pursue a highly experimental and exploratory approach to the design process, which was maintained throughout the design study. As the problem formulation stage was revisited later in the design study, the visualization researchers turned once again to visualization theory, digging into specific visualization techniques that best supported the effective encoding of the evolving data abstraction for the tasks at hand.

4.2 ADR Phase 2: Building, Intervention, Evaluation (BIE)

During the BIE stage, the team developed a broad array of technology probes to understand three core questions: What sonic devices are interesting to the poetry scholars? What are the scholars interested in doing with the sonic devices? How can visualization support their exploration of the sonic devices? Overarching these questions was the larger investigation into the role visualization could play in poetry scholarship. Sein *et al.* [5] describe small numbers of BIE cycles relating to beta and alpha prototypes of software systems in their examples of ADR. On reflection, our design study consisted of a series of BIE cycles that occurred at multiple scales: one high-level, overarching BIE cycle examining the role and impact of technology on poetry scholarship; three mid-level cycles focusing on sound, visualization design, and the development of the Poemage tool; and many rapid, low-level cycles of iteration, expansion, and refinement, each involving a planned and active intervention, evaluation of effect on the poetry scholars, and subsequent reflection to establish knowledge gained and to drive design decisions. Each scale warranted different types of evaluation, with the higher level scales incorporating more formal evaluation, and the lower level scales using quicker, lighter weight methods. These low-level BIE cycles are of particular interest for future investigations into reliability via documentation and recording of evaluation efforts. Furthermore, the different scales of BIE cycles may be particularly important in visualization as it is so often driven by discovery.

The multiscale BIE cycles for the Poemage design study are roughly depicted in Fig. 4.1. The three horizontal lines reflect contributions from different members of the Poemage team: the visualization *researchers*; the poetry scholars, or *practitioners* in ADR parlance; and *end-users* beyond the team. The top line relates to development on the part of the visualization researchers as they produced functionality for intervening in the practices of the poetry scholars. The middle line indicates an intervention as the developed artifact was deployed to the scholars. We should emphasize that this is the *crucible* of action, where the ADR team, design, and data interact in an authentic setting, and where a plausible, theory-ingrained artifact is used by a practitioner to establish knowledge in both the application and visualization domains. Further development, and reflection and learning, result from evaluation of these planned actions. The bottom line corresponds to the deployment of the Poemage tool to users beyond the team. Lam *et al.* [35] describe this as *"deployment ... in the field,"* which offers opportunities for summative evaluation, as is described

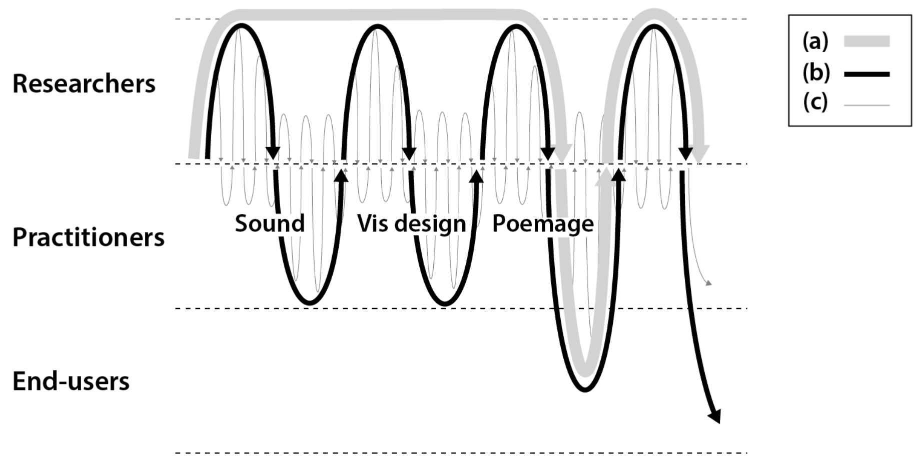

Fig. 4.1. Multiscale BIE cycles. Our BIE cycles occurred at multiple scales: (a) high-level BIE cycle focusing on the role of technology in poetry scholarship; (b) mid-level BIE cycles focusing on sound, vis design, and the development of Poemage; (c) low-level cycles involving fast, informal feedback.

in multidimensional in-depth long-term case studies (MILCs) [36]. In what follows, we outline the three mid-level BIE cycles that were core to the Poemage design study.

The first mid-level BIE cycle focused on sound and sonic devices. Via an informal survey followed by semistructured interviews, the visualization researchers worked with the poetry scholars to determine which sonic devices would be most interesting to explore in the close reading of a poem. The identified sonic devices were translated to code within an interactive system that extracted sets of words in a poem that were related via the various devices. The visualization researchers used this software as a technology probe to test the selected devices, and to understand how the poetry scholars might explore such devices within a poem. Evaluation of the technology probe ranged from casual feedback to highly structured interviews. Insights from the initial technology probe motivated the visualization experts to develop a language along with a formalism for specifying and analyzing a broad range of sonic devices — all of which the poetry scholars blanketed under an extended definition of rhyme — within a poem. This language and formalism was subsequently implemented in a system called RhymeDesign [17]. Evaluation for RhymeDesign was formal, including both case studies and a survey testing the expressivity of the RhymeDesign language against examples of interesting sonic devices collected from an extended network of poetry scholars.

The second mid-level BIE cycle focused on the design of visual representations. This cycle included explorations of different visual representations of the data abstraction as well as experimentation with different visualization and interaction techniques to support the exploratory tasks observed and identified in the previous BIE cycles. As the second BIE cycle progressed through a series of rapid, high-frequency interventions, the poetry scholars' interest evolved from browsing through sets of words detected by the system to exploring the interaction between these sets across the space of the poem. This new focus inspired the team to revisit a metaphor relating sound in poetry to flow developed previously by one of the poetry scholars, which in turn informed the visual notion of sonic topology. This cycle was guided by regular, rapid, and informal feedback from the poetry scholars on ideas and prototypes (sketches, screen captures, live demos, etc.) shared in person or remotely.

The third and final mid-level BIE cycle focused on the development of the Poemage visualization tool. During this cycle, valuable features, interactions, capabilities, and design elements were extracted from previous BIE cycles and compiled into a multilinked view system. Following an initial beta-testing deployment period in which poetry scholars from an extended network were given several weeks to experiment with incorporating Poemage into their practices, the visualization researchers conducted contextual interviews and case studies. In preparation for these focused evaluation sessions, the poetry scholars wrote experiential, qualitative narratives about their experiences using Poemage, which they discussed during the interviews. Although this preparatory writing was not asked of the poetry scholars, they expressed that this was a natural and productive method of reflection within their field, and an exercise they were inclined to complete regardless. In reflection, we note that insights like these could point to new forms of evaluation for others working in the digital humanities.

4.2.1 Adherence to P3, P4, and P5

In reflecting on the BIE cycles, we found that reciprocal shaping (**P3**) often occurred during close collaboration between the visualization researchers and poetry scholars, that evaluation (**P5**) occurred rapidly and informally during periods of intervention, and that the cycles *supported* mutual influence (**P4**) by creating a gradual decrease in separation between the knowledge states and the roles of the researchers and scholars.

One specific example of reciprocal shaping occurred around the development of a particular feature in Poemage, which came to be known as the *beautiful mess*. The beautiful mess, shown in Fig. 4.2, displays all the detected sets for a given poem, resulting in visual clutter and significant overplotting. Although this feature was explicitly requested by the poetry scholars, it was met with

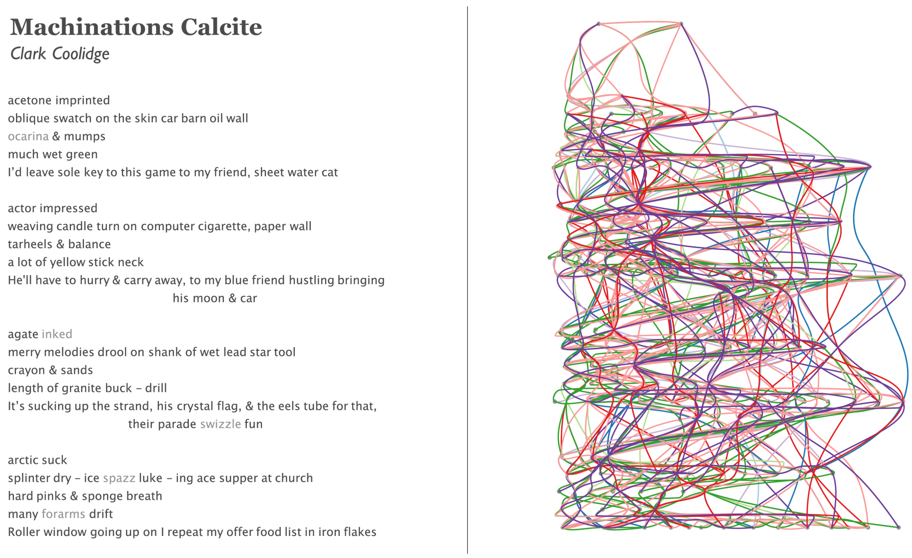

Fig. 4.2. The *beautiful mess* feature. An example of the feature applied to Clark Coolidge's "Machinations Calcite." Development of this feature exhibited elements of reciprocal shaping (**P3**) and guided emergence (**P6**).

a degree of resistance by the visualization researchers as it contravened visualization conventions that value clarity and readability [91]. The poetry scholars argued, however, that the messiness resonated deeply with them, because it captured the energy and excitement they felt during a close reading, as well as serving as a visual representation of the untangling task they confront with a new poem. Ultimately, the inclusion of the beautiful mess not only led to one of the more important insights of the work, but also helped to engage and gain the trust of the poetry scholars.

The reciprocal shaping of the beautiful mess contradicts the visualization theory brought to bear on the design process (**P2**). The design, however, worked for the poetry scholars, emphasizing the importance of reciprocal shaping and mutual influence. The beautiful mess was considered to be a strange anomaly of the design study, and precisely what to make of it remained unclear to the visualization researchers. An understanding of the notion of reciprocal shaping along with structured guidance to embrace and nurture this element of the design process and its contribution to design and knowledge acquisition, such as that provided by ADR, may have resulted in more features like the beautiful mess, and in more directed learning and evaluation around such features.

Mutual influence also played a significant role in this research. Each team member contributed a different level of expertise in her own field, a different level of expertise in the other domain, and a different level of openness to deviating from theory and convention. Throughout the design process, the poetry scholars developed a computational way of thinking about their scholarship, which they discussed and reflected on in multiple articles and talks to the humanities and digital humanities communities [69], [87]. This, along with insights gained throughout the collaboration, led one collaborator to develop new theoretical thinking about the relationship between human and machine in the context of the digital humanities. On the other end of the collaboration, the visualization researchers learned to embrace the poets' broad and imprecise definition of rhyme and developed an openness to deviating from conventional visualization methods and principles. Additionally, the visualization researchers learned to incorporate a more extemporaneous element into their research — one that reflected the nature of their collaborators' poetry scholarship. Furthermore, revisiting a close reading of a particular poem, and the particular analysis that led to a new interpretation or insight, was a regular tactic used by the poetry scholars to illustrate a point. Thus the visualization researchers had to develop enough of an understanding of poetry, poetry analysis, and close reading in order to interpret the point being made, and translate it to the space of visualization research.

Lastly, authentic evaluation played an integral role in shaping the research and design process. Fast and informal feedback guided the research team toward pursuing sonic devices and facilitated the design process. At various points throughout the first and second BIE cycles, the visualization researchers sat with the poetry scholars and iteratively tested and evaluated new features, interactions, and visual encodings. In addition, consistent feedback helped the visualization researchers identify and build on elements of the research process that engaged the poetry scholars, increased their trust in the technology, and were disruptive in some interesting sort of way. In retrospect, recording and reporting these kinds of findings in a more structured and perhaps comprehensive fashion, as ADR begins to facilitate, would have increased the reliability of the design process.

The evaluation strategies, particularly as they applied to interviewing techniques, evolved and shifted throughout the BIE cycles based on feedback and reflection. For example, the first round of interviews was highly structured, but it became clear that semistructured interviews were much more appropriate since the poetry scholars needed very little prompting and came to the interviews with valuable insights that would have been hard to elicit via preconceived questions. As another example, elements of poetry scholarship found their way into evaluation tactics. The primary example of this was the experiential, qualitative narratives written by the poetry scholars that were incorporated into the evaluation of Poemage. Thus, mutual influence and reciprocal shaping had an effect on *evaluation* as well as on design. This would not be welcome in the kinds of isolated objective evaluation that lab studies permit, but a reflective methodology such as ADR provides a means for such flexibility while providing reassurance regarding reliability in applied work. Had this type of mutual influence and reciprocal shaping occurred earlier in the design study, or had the team been following a methodology that explicitly encouraged this awareness and flexibility, the team may have sought and benefitted from more opportunities of this kind.

4.3 ADR Phase 3: Reflection and Learning

Throughout the design process, the visualization researchers reflected in order to shift and shape the direction of the project, operationalize poorly defined tasks, and extract insights. For example, during the first BIE cycle, it became clear that the poetry scholars embraced a broad and imprecise definition of sonic similarity, motivating the visualization researchers to move beyond straightforward rhyme detection. The result was the development of a formalism for describing sonic similarity computationally, and the implementation of the RhymeDesign tool. Another reflective moment occurred when the visualization researchers observed a spectrum of ways in which the poetry scholars were using the Poemage tool, leading to a realization that one role of technology in poetry scholarship is for creativity support, as opposed to data analysis.

4.3.1 Adherence to P6

Moments of guided emergence occurred throughout the design process. An illustration of this is the beautiful mess example described in Section 4.2. The visualization researchers were guided by conventions surrounding clarity and readability, and they initially resisted even experimenting with the feature. As the poetry scholars continued to push for the feature, however, one of the visualization researchers became more receptive. As mutual influence was established through validation of the technique, the other visualization researcher was eventually persuaded. In hindsight, this experience taught the visualization researchers to be more open to precisely the notion of guided emergence. At the time, the precise impact and takeaway of this anecdote remained unclear to the visualization researchers; however, the lesson was presented in the visualization publication about the Poemage design study as a kind of guideline that encouraged others to adopt the same openness in their research. **P6** directly confirms and articulates the importance of such experiences to the design process, and gives weight to any associated lessons and formulated guidelines.

4.4 ADR Phase 4: Formalization of Learning

Formalization of learning in this project occurred, by and large, during the writing phase of the research. During this period, the visualization researchers looked back through the entire project, gathering and formalizing the elements of the project that had potential for benefiting the visualization community as a whole. Some formalization came out of the problem characterization and data abstraction, as is typical in the reporting of design work in visualization research [92]. Other formalization came out of reflecting on the project as a whole, including insights surrounding creativity support tools and conducting design research in the digital humanities. Additionally, the visualization researchers revisited the most interesting challenges encountered throughout the research, especially those surrounding evaluation and appropriate measures of success, and formalized them into open research questions for future work. A desire to formalize learning surrounding the reciprocal shaping and the disruption occurred throughout the design process, but the lack of guidance and language for doing this in existing methods for visualization design left the visualization researchers with little confidence in such an endeavor.

4.4.1 Adherence to P7

Although the results of this research were highly specific and designed to meet the interests and needs of a very small group of poetry scholars, the visualization researchers generalized elements of the process and design to various levels of abstraction. For example, the poetry scholars' interest in exploring the role of technology in their scholarship practices motivated some speculation about possible implications in the arts and in other fields that value novel interpretations and creative thinking. At a much lower level, whereas Poemage was designed to support a very specific research activity — the close reading of American English free verse poetry — formalizing the data abstraction allowed the visualization researchers to speculate about possible applications to other set visualization problems. In retrospect, we wonder whether taking an ADR approach might have facilitated framing these outcomes more effectively, consistently, and ultimately more reliably.

4.5 Discussion

Applying ADR retrospectively to the Poemage design study revealed a number of significant moments and insights that we struggled to articulate using existing models for conducting and reporting visualization research. ADR provided structure and organization for analyzing the impact of the human-centered and disruptive elements on the process and design, as well as the impact of the collaboration on the learning that occurred in both domains: visualization and poetry.

In retrospect, we find that our key visualization insights from the Poemage design study, described in Chapter 3, can be expressed in terms of ADR's principles and stages. The importance of *really* listening to one's domain collaborators, and the benefits of embracing and leveraging challenge and disruption, can be reframed as an adherence to ADR's principles of reciprocal shaping, mutually influential roles, and guided emergence. The question of how to reliably conduct, validate, and report on these collaborative, human-centered elements of applied visualization research, along with our insight about approaching visualization systems as dynamic technology probes, can be reframed in terms of ADR's stages of building, intervention, and evaluation; reflection and learning; and formalization of learning.

Even though we are unable to draw conclusions around ADR's utility as a guiding methodology, our retrospective application of the framework leads us to hypothesize that incorporating elements of ADR into future design studies will enable better navigation and evaluation of the design process, as well as the facilitation of new kinds of learning.

CHAPTER 5

DESIGN STUDY: A FRAMEWORK FOR EXTERNALIZING IMPLICIT ERROR USING VISUALIZATION

The third piece of work in this dissertation is a design study with global health experts. We viewed this design study as an opportunity to test and develop ADR concepts in the wild, and to explore new approaches to capturing and reporting on the design study process, an element that was lacking from the ADR framework. We adopted a practice of taking field notes to facilitate reflection and learning and adherence to ADR principles, and to experiment with approaches to documenting the research process. We were deliberately more attentive to the backgrounds and perspectives of our domain collaborators as well as to our own incoming perspectives and assumptions. We took an action research approach, welcoming disruption on both sides of the collaboration and employing visualization primarily as a tool to learn with.

These actions shaped our research process and results. The primary contribution of the design study is a theoretical framework that is rooted in domain experts' background knowledge and perspectives. The framework evolved over multiple cycles of intervention with global health experts, combined with critical reflection and formalization of learning by visualization researchers. The framework signifies disruption on three levels: first, the framework seeks to disrupt the global health experts' views of and relationships with data and visualization; second, the framework is rooted in a disruption to the visualization researchers' thinking about the role and value of visualization; and third, the framework itself is a disruption to broader visualization thinking about error and uncertainty.

ADR also influenced the way we presented this design study to the visualization community. ADR gave us confidence in our framework as a valuable research contribution, and confidence to deviate from our traditional tool-centric reporting methods. In an effort to increase reliability and transferability of our research, which we saw as an evolutionary step beyond ADR, we include a rich, reflective, and verifiable process description, combined with an interactive timeline of our field notes as supplemental materials. Adapting the language of ADR, we frame our process description around cycles of intervention followed by a period of formalization of learning. For each stage, we report on the artifacts that were generated and reflect on the learning that occurred.

In what follows, we present the details of our design study. We then reflect more deeply on our experience applying and extending ADR concepts in the wild.

5.1 Overview

The research we report on in this chapter stems from a 6-month field study at the United States Agency for International Development (USAID) within the Bureau for Global Health. During this study, we collaborated with global health experts working to combat the Zika virus and associated health threats in Latin America and the Caribbean. The collaboration displayed the characteristics of a classic design study [8]: there were data, there were clear domain tasks, and our collaborators were interested in exploring new approaches to visualization. By the end of the field study, we had developed an interactive visualization tool for analyzing Zika data; positive feedback from stakeholders attested to its usefulness.

Despite this success, however, we noticed a hesitation by our collaborators to embrace the new tool for their analysis. In probing their reluctance, we confirmed that although the tool was a good reflection of the Zika outbreak data, the data were not an accurate reflection of what the experts knew to be true about the outbreak in the region. We pivoted to focus on this problem, and discovered that the distributed, heterogeneous nature of generating and aggregating data about the outbreak within and across multiple countries resulted in inherently erroneous data. Even though the data did not reflect these errors, the experts had a wealth of domain knowledge about their existence, their impact, and their source.

We use the term *implicit error* to describe these data discrepancies. Implicit error is a type of measurement error that is inherent to a dataset but not explicitly recorded, yet it is accounted for qualitatively by experts during analysis, based on their implicit domain knowledge. We developed a description of implicit error based on our analysis of data discrepancies in Zika outbreak data — we speculate that our description is relevant to implicit error in a variety of domains — and we explored annotation as a mechanism for externalizing and analyzing implicit error using visualization. This work points to the potential of externalized implicit error for supporting more effective data analysis, for transferring insight between experts, for serving as a memory of institutional knowledge, and for enabling modeling and abatement of systematic error in data.

Grounded in our design study with global health experts, the first contribution of this work is a framework for reasoning about and externalizing implicit error using visualization. The framework includes a description of components of implicit error that are important for downstream analysis, and a process model that details the role of visualization in externalizing and analyzing implicit error. We demonstrate the framework in practice through a visualization tool designed to support externalization of implicit error in Zika outbreak data.

The second contribution of this work is an extensive description of our 18-month research process, supported by a practice of taking frequent field notes, which we propose as a rich, reflective, and verifiable exemplar summary of design study research. Through this process description, we hope to contribute to the ongoing dialog within the visualization community surrounding the recording and reporting of applied research process and findings.

5.2 Related Work

The specific designs of the tools developed throughout our research draw from previous work focusing on the design and development of decision-support and surveillance tools in the context of public health. Our work primarily fills a gap between the broad span of literature that acknowledges the prevalence of implicit error, as we have defined it in this work, across a variety of domains, and existing visualization work that models the externalization of domain knowledge.

5.2.1 Public Health Decision Support and Surveillance Tools

A considerable span of research focuses on developing visualization and visual analytics systems to support decision-making and surveillance for epidemics and other public health emergencies [93], [94]. Public health decision-support tools typically avoid the issue of implicit error by employing epidemic models to simulate the evolution of an outbreak and the impact of various responses [95]–[99]. This approach is employed in some areas of global health, but we found in our field study that much of the analysis by global health experts operates on raw, epidemiological surveillance data.

A number of public health surveillance tools facilitate exploration and analysis of raw surveillance data for real- and near-real-time outbreak detection, particularly in the context of biosurveillance [100]–[105]. These tools emphasize the important role of situational awareness, which is the perceptual understanding of the context in which a situation takes place [106], in appropriately interpreting surveillance data, and provide support by incorporating contextual information, such as relevant current events and unofficial outbreak data, to reflect domain experts' mental models of situational awareness. In addition, these tools and a subset of public health decision support tools [96], [107] explicitly rely on users to juxtapose the presented data with their own domain knowledge during interpretation and decision-making. As we found in our field study, implicit error is predominant in contextual expert knowledge, and thus supporting situational awareness inherently supports incorporating this knowledge into analysis. In this work, we propose a more direct approach to incorporating and compiling this subset of contextual expert knowledge.

5.2.2 Implicit Error and Knowledge Externalization

The existence of implicit error is well established within public health surveillance [108]–[111] and across domains in which humans and societies play a central role in data acquisition, curation, and interpretation [112]. Such domains range from emergency response and disaster operations management [113], [114] to various forms of risk assessment [115], healthcare, and medicine [116]. In public health, surveillance data are often published along with disclaimers such as the following: "case numbers are generally a poor indication of the true burden of disease. To interpret these numbers, one needs to consider both epidemiological patterns and data collection efforts in specific" countries [117]. Attempts to standardize data generation pipelines are a primary approach to minimizing these errors across systems [118]. These efforts are strengthened by methods for evaluating data quality and compliance [119]. There is a general acknowledgement, however, that errors will persist despite these efforts [112], [120]. In this work, we formalize the notion of implicit error, and propose a framework to support externalizing implicit error by domain experts through visualization.

There is also work that explicates the importance of context, history, background, and knowledge, described as "the stuff around the edges" [121] in accurately interpreting a piece of information. This work warns that "attending too closely to information overlooks the social context that helps people understand what that information might mean and why it matters." In this chapter, we attempt to directly capture and explicate the *stuff around the edges* in order to shed light on unaccounted for errors within data, extending the known benefits of employing contextual knowledge to enhance recall and comprehension in visual analysis [122]–[125].

Visualization is widely recognized as a platform for facilitating the projection of contextual domain knowledge onto data. This facilitation is captured in visualization models as a key component for meaning-making and insight generation [43], [126]. Knowledge-assisted visualization models explicate the externalization of expert knowledge into a computational representation that can be used to drive system specifications and simulated cognitive processes [126]–[128]. The work presented in this paper builds on these models, articulating the role of information and annotation in the externalization of domain knowledge, as well as in transferring knowledge across experts.

An entire subfield of visualization focuses on the visual representation of error and uncertainty [129]–[132], and a large body of work within this subfield focuses on visualizing the errors and uncertainty of geographic data and the associated data attributes [133]. This uncertainty visualization work, however, focuses on quantifiable measures of error and uncertainty, with some attention to categorical measures [134]. Although implicit error stems from the same sources as quantifiable measures of error, and has the same impact on reported values, its qualitative nature requires a different set of considerations and visualization approaches, which we explore in this work.

Finally, work in data provenance focuses on capturing the nuances of data generation and processing pipelines [135]–[137], for example in areas such as human terrain visual analytics [138]. Additionally, *insight provenance* supports externalization of implicit knowledge about the data, primarily through annotation. Our work could enhance these fields by providing an explicit mechanism, the externalization of implicit error, for capturing insights about potential sources of error.

5.3 Problem Domain Background

In early 2016, the Zika virus and associated neurological disorders such as microcephaly were declared a public health emergency of international concern. Since then, global health experts have worked to plan and implement effective response efforts. These efforts involve understanding the risk and impact of Zika within and across countries and regions around the globe, and distributing resources and interventions accordingly. Experts working to assess and suppress the spread of diseases like Zika rely on two sources of information: *outbreak data* that track the spread of the virus across a region, coupled with information about the demography and geography of the region; and *response data* that describe international response efforts underway. Using these data, experts seek to understand how an outbreak is spreading across regions, assess the risk and relative impact of the outbreak on underlying populations, and understand how these risk and impact factors change over time. Doing this involves identifying *hotspots*, which are heavily impacted regions, and *coldspots*, which are lightly impacted or unaffected regions, and predicting future locations of each. Once these hotspots and coldspots are identified, outbreak data are compared against response data to assess the appropriateness of response efforts.

A key component of outbreak data is *epidemiological surveillance data* (epi-data), which track reported cases of a disease and associated health issues through a systematic process of collection, analysis, and dissemination [139]. In the case of the Zika virus, epi-data, which are reported weekly, include counts of both suspected and confirmed cases of infection along with counts of other related issues such as microcephaly and Guillain-Barre Syndrome; the set of reported attributes are referred to as disease *indicators*. Additionally, epi-data are often augmented with proxy measures for a disease such as, in the case of Zika, data on certain mosquito populations (*entomological surveillance data* or ento-data), as well as epi-data of other related mosquito-transmitted diseases such as dengue, which has been tracked for years. Ento-data were not included in the present study because the focus at the time of tool development was on publicly available case reports. Epi-data are further augmented with geographic data such as rainfall amounts and characterizations of low-lying regions, as well as with demographic data such as population density distribution and poverty levels. For global health experts, epi-data convey the impact of an outbreak, whereas geographic, demographic, and ento-data help to convey its risk.

Due to the borderless nature of outbreaks, the collection, analysis, and dissemination of epi-data are conducted by a hierarchy of organizations. At the finest resolution, measurements of disease indicators are collected by local clinics and governing subcountry health offices. These data are reported to a country level Ministry of Health office that compiles and releases data reports regularly, usually as PDFs containing numerical data tables along with related charts, choropleths, and text. We note that although this is the established best practice, the consistency and degree to which epi-data reports are published varies from country to country. From here, the regional arm of the World Health Organization (WHO), an agency of the United Nations specializing in international public health, works with ministries of health to collect reports, which it then compiles into a weekly regional report, made available as a raw table or as a table in a PDF.

In this work, we collaborated with global health experts working to combat the Zika virus in

Latin America and the Caribbean. Our collaborators have interdisciplinary backgrounds in public health combined with epidemiology and a range of social and biological sciences. In addition to this background, their expert domain knowledge includes an in-depth understanding of the countries and regions that they individually serve: from the nature and strength of the epidemiological surveillance systems, to the political, economic, cultural and geographic contexts. This regional domain knowledge plays a critical role in assessing the impact and risk of a transnational outbreak, as well as in developing and refining effective response efforts.

5.4 Process, Artifacts, and Reflection

In this section, we report on the core phases of our 18-month design study using a rich description of the methods we used and the artifacts we created, combined with reflective syntheses of what we learned along the way. During the first 6 months, we conducted a field study in Washington, D.C. at USAID's Bureau for Global Health. The field study began with a preconditioning phase [8] during which we interacted with a variety of teams, developed an understanding of the data analysis needs and challenges across the Bureau, and established relationships with a range of domain experts and other stakeholders, who would later provide invaluable feedback on the broader applicability of our research findings. Furthermore, through presentations and visualization design work, we established credibility with various stakeholders, which helped us to obtain the necessary buy-in to pursue design study research without the guarantee of deliverables. We winnowed our efforts to a collaboration with global health experts working to combat Zika, and focused the remainder of the design study on their analysis needs. The field study was proceeded by 12 months of research conducted from the University of Utah. To protect privacy, a number of low-level details about participants and the organization have been omitted from this section.

We used this project as an opportunity to investigate new ways of approaching, recording, and reporting design study research. We viewed each design and development phase as an opportunity not only to build a deployable tool, but also to use the tool itself to probe the problem space and learn more about the challenges faced by our collaborators. In support of this learning, we decided early on to capture notes and insights as frequently as possible, both for our own reflective analysis and for auditing by others for validation. We adopted a practice of taking field notes following meaningful interactions, providing a log through which we could trace the development of insights and ideas. Although many of the details captured within these field notes are confidential, we provide an interactive timeline of high-level field note summaries in supplemental materials. We report on the project using a rich description of our process, with an eye toward articulating the moments and artifacts that we believe were central to building and shaping the research results. We put forth the extensive process description, along with our practice of taking and releasing field notes, as an exemplar contribution toward the ongoing inquiry into ways of increasing the validity and transferability of design study research.

5.4.1 Learning About Zika Outbreak Analysis

After 2 months of preconditioning, we began the process of deeper collaboration and iterative design work with the domain experts. Our goals during this phase were to establish an understanding of the domain problem, described in Section 5.3; develop an abstraction of the underlying data and tasks; and design a visualization system to support our collaborators' analysis. This phase spanned the last 4 months of the field study.

During this phase, we collaborated with nine domain experts. Our two primary collaborators were global health experts who helped us understand the domain problem and the associated data and tasks, as well as how Zika experts interact with, and interpret, outbreak and response data. We also collaborated with three fellow tool builders from the USAID's in-house resource for spatial analysis and geographic information system (GIS). Prior to the start of the field study, our collaborators had reached out to members of this resource for visualization support, working with them to formulate high-level tasks and to begin the process of compiling and visualizing the relevant data in ArcGIS — important preconditions for design study. These fellow tool builders agreed to let us take the lead on the project, helped us establish our understanding of the challenges surrounding the Zika data and tasks, and also provided a valuable resource for brainstorming, triangulating and validating ideas, and gathering feedback on prototypes.

Additionally, we worked with four tertiary collaborators who deal first-hand with challenges around the collection, processing, and analysis of global health data. These tertiary collaborators were experts both in data processing and evaluation, as well as in a range of global health efforts. They provided another valuable resource for brainstorming, triangulating ideas, and gathering feedback on prototypes, and also provided insights on the broader applicability of our research findings across global health.

Throughout this phase, we conducted informal interviews with all collaborators, meeting monthly to biweekly with our primary collaborators, bimonthly with our fellow tool builders, weekly with one of our tertiary collaborators, and occasionally with other tertiary collaborators. We additionally conducted a think-aloud with a primary collaborator using the existing ArcGIS platform developed by our fellow tool builders. In nearly all cases, these meetings were recorded and *reflectively transcribed*, a process of reflection and note-taking while listening to an audio recording, seeking to capture the gist of discussions along with insights acquired during the transcription. We stored these reflective transcriptions as field notes. We took additional field notes both before meetings to outline goals and assumptions and after meetings to capture initial reactions and insights.

After a set of initial interviews, during which we also worked with our collaborators to gather relevant data, we began working on the design of a technology probe for data and tasks using a rapid prototyping approach. The data and tasks technology probe allowed us to probe the analysis needs of our collaborators and the nuances of the data and helped us build our understanding of the problem space more generally [31]. The final design of the probe reflected our understanding of the problem, including the data and task abstraction, at the conclusion of this phase. Prototyping began with hand-drawn sketches, Adobe Illustrator mockups, and low-fidelity sketches, and then proceeded to a high-fidelity visualization tool implemented in D3.js. The rounds of prototyping were interspersed with feedback sessions with our primary collaborators, which guided further refinements to the overall design. The feedback gathered around the design and use of the data and tasks probe triggered new insights and hypotheses for us surrounding visualization research opportunities and ways in which our work could benefit both our collaborators and the larger global health community.

Further validation of the data and tasks probe — and thus, validation of our data and task abstraction for the problem — was obtained through presentations of the probe to a broader set of stakeholders. We presented the probe to a larger group of Zika experts over a teleconference, as well as in person to other global health experts, and ultimately, by invitation, to a larger group of stakeholders.

5.4.1.1 Artifact: Data and Task Abstraction

Two primary sources provide epi-data on the Zika virus to the international public health community: the Ministry of Health (MoH) offices of individual participating countries and the World Health Organization (WHO). The MoH data include country and subcountry level epi-data for each participating country. These data report on a set of indicators that varies both from country to country and between the two resolutions of the data. Differences in reported indicators are due to variation in what individual countries deem important to measure and report. The WHO data, compiled from both public and private sources of MoH country level data, include country and regional level epi-data. The sets of indicators reported for these two resolutions of the data are consistent across countries, but differ from the sets of indicators reported in the MoH data. These differences reflect what the WHO deems important for monitoring outbreaks across countries and most useful to global health organizations. The consistency and reliability of the WHO data make it a primary source of information for global health experts. The aggregation of these data, however, makes them particularly prone to discrepancies resulting from variations in countries' surveillance systems, an issue that global health experts are aware of. Thus, the MOH data, with their finer resolution, promote understanding of low-level trends of a disease outbreak.

Both the MoH and the WHO data are reported on a weekly basis during the height of an outbreak. The data are thus provisional; they reflect a snapshot of known epi-data at a particular moment in time. One consequence of the provisional nature of the data is that retrospective updates to these data are published downstream, leading to temporal data discrepancies. Examples of this are falsely confirmed cases or local cases later found to be imported from other countries. Within the WHO dataset, these discrepancies are published as footnotes alongside the indicator values.

The epi-data are augmented with two types of metadata. Demographic and geographic metadata capture statistics surrounding poverty, population density, and rainfall. These data are reported at both the subcountry and country level by various publicly available and established databases. Response metadata are reported at the country level, with some finer level data at the subcountry level for a subset of response programs. The response metadata include information about the line of response effort, such as mosquito population control or health services, the partnering organization, and the target geographic area and population. The metadata report on current conditions and efforts. Whereas the epi-data are the core data used to characterize the spread of the Zika virus, the metadata are used to summarize response coverage and assess the risk and impact of the disease on underlying populations. Although beyond the scope of the current work, future plans to further augment the analysis with ento-data will help vector-borne disease specialists predict future cases and the spread of the disease.

The primary tasks of our collaborators are threefold. First, they need to identify and characterize how the Zika outbreak is evolving based on indicators of the disease over time and space. Second, they need to identify and characterize the outbreak's impact on, and risk to, underlying populations based on the geographic and demographic metadata. Third, they must assess whether the response coverage is appropriate with respect to the evolving outbreak and its impact on and risk to underlying populations while also considering factors like equity. To give an example, suppose a global health expert is looking at an epi-data indicator that reports cumulative confirmed Zika cases across a country. She identifies a part of the country with a relatively high number of cases as a hotspot. Looking at the demographic metadata, she sees that the hotspot is in a densely populated area with high poverty levels. This is not surprising, given her knowledge of the disease and associated risk factors. She reviews the response data in that part of the country and finds a number of different partnering organizations working there, covering all lines of effort. She confirms that the response is appropriate, as everything possible is being done to combat the Zika virus in that area. She thus recommends no reallocation of resources.

5.4.1.2 Artifact: Data and Tasks Technology Probe

The data and tasks probe developed during this phase, and shown at the top of Fig. 5.1, was designed to represent and support the data and task abstraction that we developed based on our collaborators' analysis needs. Developed for the web using the D3 and Leaflet Javascript libraries, the probe uses a standard linked-view approach to explore geospatiotemporal data, with customizations to support specific requirements of Zika experts. The probe supports exploration and comparison of outbreak and response data at two levels of resolution: the regional level, showing WHO data for the region and associated countries; and the country level, showing MoH data for a country and its subcountry areas.

At each level, line charts displayed in a chart view allow users to explore trends in different indicators over time and compare these trends against the associated geographic and demographic metadata. The chart view is linked to a map view showing a temporal snapshot of epi-data encoded as a choropleth. Response data are overlaid on the choropleth, either as glyphs at the regional level or as textured shapefiles at the country level. At the country level, users can additionally view epi-data over time as small multiples of choropleths.

As a final addition, the probe also supports lightweight annotation, meant to probe the potential of the mechanism for capturing implicit domain knowledge about the data. This technology probe informed design recommendations for a Tableau-based tool, under development by global health

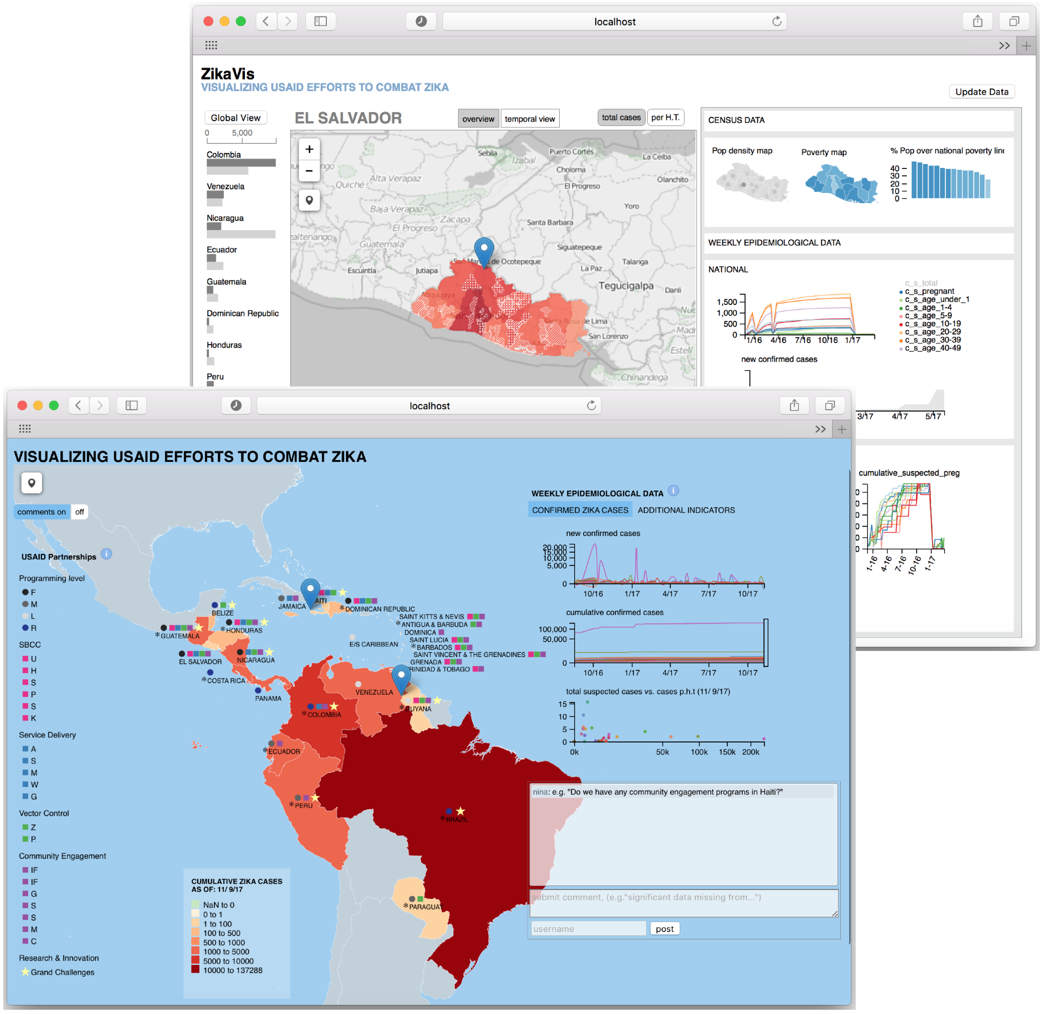

Fig. 5.1. Technology probes. Developed during the second (top) and third (bottom) phases of the project, both versions allow users to explore country and subcountry data from countries' MoH offices (top), and country level data from WHO (bottom). The third phase probe integrates a fully implemented annotation platform.

experts to support sustainability and continued development of visualization for Zika outbreak analysis.

5.4.1.3 Reflection

Regular feedback and triangulation from our primary collaborators, fellow tool builders, and tertiary collaborators provided incremental validation of our understanding of the problem, and heavily shaped the design of the data and tasks technology probe. In addition, the positive feedback we received during presentations of the probe served as informal validation that our results are more broadly relevant to global health beyond Zika outbreak analysis.

More interesting, however, was feedback on the probe from the larger group of Zika experts confirming our growing suspicion that although the probe was an effective reflection of the data, the data were not an accurate reflection of what the experts knew about the current status of the Zika outbreak. As one of our collaborators put it, testing the probe required “suspending disbelief” around the quality, consistency, and availability of the data.

What also became increasingly clear during this phase, and was subsequently confirmed by the larger group of Zika experts, was that knowledge about discrepancies in the data exists largely within the minds of Zika experts. The first indication of this implicit knowledge emerged during a feedback session featuring the probes’ regional level choropleth displaying cumulative confirmed cases of Zika on a per country basis. Brazil was displayed in dark red whereas Colombia appeared as a lighter orange. One collaborator noted that whereas Brazil reports all cases, Colombia runs a full investigation prior to making any reports. The implication of this comment was that visualization of the official data was indicating a relationship between the countries that conflicted with our collaborators’ understanding of the outbreak.

We witnessed similar data qualifications on a number of other occasions. As we probed deeper, we came to understand that our collaborators’ regional domain knowledge — their in-depth understanding of regional context and response efforts — included an extensive mental database of the idiosyncrasies that go unaccounted for in the data generation pipeline, and that lead to errors in the official data reported for a region. Our collaborators learn to view data and data visualization through the lens of this contextual knowledge and, furthermore, assume the presence of errors when viewing data and visualization from outside their own region of expertise. In the Brazil-Colombia example, our collaborator was mentally adjusting the colors of the two countries in order to better account for the discrepancy in the data. When we asked another collaborator about this in a follow-up conversation, she responded with “Yeah, you kind of have to.” Our suspicion that the cognitive load required to make these mental adjustments decreased the potential impact of visualization tools for our collaborators led us to reconsider our goals for the project, and to pivot toward tackling the upstream problem of discrepancies in the Zika epi-data.

5.4.2 Learning About Discrepancies

We pivoted to focus on understanding and characterizing discrepancies in Zika epi-data, and on investigating the potential of annotation as a mechanism for externalizing expert domain knowledge about data discrepancies. To meet these goals, we extended the data and tasks probe developed in the previous phase to include full annotation support. We evaluated the new annotation probe though a workshop with a larger group of Zika experts. Other than the workshop, this phase was conducted at the University of Utah and spanned approximately 2 months. During this phase, we also began working with an additional primary collaborator — an institutional contractor working full-time on the Zika response. This collaborator was heavily involved in the remainder of the study and is a coauthor on the publication of the results of this design study.

The workshop provided an opportunity to meet face-to-face with, and gather feedback from, Zika experts based in countries across Latin America and the Caribbean. The workshop was held in a computer lab equipped with Windows desktops, and lasted 1 hour and 45 minutes. One visualization researcher facilitated the workshop and 13 Zika experts participated. The workshop began with a brief presentation reintroducing the project and demoing the annotation probe. The presentation was casual and interspersed with discussion. It was followed by two hands-on activities bookended by group discussions.

In the first activity, participants were asked to explore the probe on the lab computers and to submit annotations using two separate features: an annotation feature for dropping and annotating pins on a map and a commenting feature for posting annotations to a message board. We provided minimal guidance on the kinds of annotations that we were looking for; however, we emphasized that we were less interested in notifications of missing or outdated data, and more interested in the nuances surrounding response efforts, specific geographic areas, populations, and recording and reporting mechanisms. Participants were given roughly 15 minutes to make submissions. This was followed by a group discussion guided by questions including: “What inspired you to submit an annotation?”; and “How would you hope someone else might use these annotations to help them interpret the data?” This discussion was followed by a second activity, in which participants were encouraged to explore the full set of annotations submitted in the first activity, followed by a discussion guided by questions including: “Were some annotations more useful or informative than others?”; and “How did the annotations impact your interpretation of the data?” Based on the activities and discussions, we collected 54 sample annotations.

Participants were provided with surveys for assessing usability, such as likes, dislikes, suggestions, etc. We concluded the workshop with another survey containing identical questions to those posed in the group discussions for the two core activities. This survey presented an opportunity for participants to reiterate thoughts and include new ideas that did not make it into the group discussions. Lastly, we conducted a short follow-up poll in an attempt to capture initial reactions about the annotation platform and get a definitive sense of whether we were heading in a valuable direction. Poll results were submitted for 10/13 participants, and are summarized in Table 5.1.

5.4.2.1 Artifact: Annotation Technology Probe

The annotation technology probe, shown at the bottom of Fig. 5.1, retains the basic functionality of the previous probe, with the addition of a fully implemented annotation platform. This platform supports generating annotations at varying degrees of specificity — from landmarks, to geographic areas, to general annotations — and at the regional, country, and subcountry levels. The probe also includes a refined set of visualizations based on feedback on the underlying design received in the previous phase.

5.4.2.2 Reflection

The feedback and data collected from the workshop gave us confidence that our focus on data discrepancies and on annotation as a mechanism for externalization was well directed. For example, one participant provided concise validation of this direction in her poll response: "A comprehensive combination of the annotation and comments features (at country and regional levels), especially with some basic, high-level coding scheme (related to programming? data quality? other?) would be incredibly useful for the global health community." Additionally, the set of annotations collected during the workshop formed the basis for our understanding and characterization of epi-data discrepancy in the proceeding phase of research.

Table 5.1. Results from the workshop poll. The purpose of the poll was to capture initial reactions by the participants about the annotation platform. Although informal, the results provided positive feedback on our proposed use of annotation as a mechanism for externalizing knowledge of data discrepancies in the Zika epi-data.

Question	**Response**
q1: On a scale from 0 to 10, how useful do you think the annotation (i.e. dropping pins) feature could be?	7.9 (avg.)
q2: Is this feature/concept worth pursuing? (Y/N)	Y (100%)
q3: On a scale from 0 to 10, how useful do you think the commenting feature could be?	7.7 (avg.)
q4: Is this feature/concept worth pursuing? (Y/N)	Y (100%)

5.4.3 Formalization of Learning

The final phase focused on synthesizing and formalizing learning from previous phases [2] surrounding the notion of data discrepancy. This phase took place at the University of Utah and spanned 5 months. During this time we employed two core methods. First, we performed qualitative analysis on a collection of descriptions of data discrepancies compiled from various sources throughout the project. Second, we engaged in a critically reflective practice to synthesize our experiences across the project. Our synthesis was informed by feedback from our collaborators and grounded in the relevant literature.

The qualitative analysis of discrepancy descriptions involved two rounds of affinity diagramming, conducted by the two visualization researchers. In the first round, the 54 annotations collected during the annotation workshop were clustered into three groups and four subgroups. The three major annotation groupings were about response data, outbreak data, and general questions and comments. Subgroupings of the response and outbreak data included updates and corrections, flagging of discrepancies in epi-data, suggestions for supplemental or higher quality data, and contextual narrative. We ultimately culled annotations about response data as well as those about questions and comments.

The culled subset of 27 annotations was then combined with six descriptions of discrepancies captured from interviews and discussions throughout the study, along with 240 footnotes published alongside regional level WHO epi-data, as described in Section 5.4.1.1. This larger set of descriptions was then used in a second round of affinity diagramming, conducted by two the visualization researchers. Major groupings included discrepancies due to inconsistencies, discrepancies due to missing data, temporal discrepancies, and contextual narrative providing higher resolution information. An example of this last grouping is "department x has a low incidence rate, since the department is mostly highland and so mosquitoes aren't endemic." Although compelling and potentially valuable, we decided to cull narrative-style examples as they extended beyond our evolving notion of discrepancy. The remaining groupings formed the basis of our understanding and characterization of epi-data discrepancies, and of data discrepancies more broadly.

To further develop, synthesize, and formalize our learning, we used an approach of *critically reflective practice*, which brings together experience, reflection, and critical thinking in an iterative process of synthesis and action in order to generate insights from experience [140], [141]. Using this approach, we reflected across the entire study, reexamining field notes, outcomes, and insights in light of the results of our qualitative analysis and our current understanding of data discrepancy. In addition, we studied existing literature across domains on relevant topics, including knowledge externalization [126]–[128], [142], [143] and sociotechnical systems [144], and emergent concepts such as gray literature [145] and systemic bias [146].

This reflection, combined with multiple rounds of writing, diagramming, and collaborative refinement of documents, resulted in the proposed visualization framework for reasoning about and externalizing data discrepancies — which we describe as implicit error — within epi-data, and potentially for implicit error in other kinds of data as well. We present the framework, the primary artifact of this final phase of research, in Section 5.5, and describe an instantiation of the framework as a visualization tool for Zika outbreak analysis in Section 5.6. Our reflective synthesis of this phase is discussed in Section 5.7.

5.5 Framework For Externalizing Implicit Error

The primary contribution of this work is a visualization framework for reasoning about and externalizing knowledge of data discrepancies, which we refer to as *implicit error*. The framework consists of a description of implicit error components that are important for downstream analysis, and a process model for externalizing and analyzing implicit error using visualization. All aspects of the framework were inspired by, and are grounded in, our collaboration with Zika experts. Reflections from this collaboration are used throughout the section to illustrate the framework concepts.

5.5.1 Describing Implicit Error

As we discovered over the course of our collaboration with Zika experts, differences in what the epi-data reported and what the domain experts knew to be true prevented meaningful visual analysis of the data. Measurement error in the data — the difference between the number of reported cases and the actual number of cases as they exist in the world — stems from the distributed, heterogeneous data generation pipeline. Differences in how cases are detected, recorded, collected, processed, and reported exist both within countries and between them. These differences are due to the embedding of these pipelines within countries' political, economic, cultural, geographic, and demographic contexts, all of which influence how various stages of the pipeline are implemented [112]. These differences accumulate as data are repeatedly compiled and aggregated, leading to inherently erroneous data. We speculate that other domains with distributed, heterogeneous data generation pipelines feature similar errors as well.

Although a precise quantification of these errors is infeasible, global health experts have extensive domain knowledge about their existence and source. We thus use the term **implicit error** to describe measurement error that is inherent to a given dataset, assumed to be present and prevalent, but not explicitly defined or accounted for. Instead, implicit error largely exists as tacit knowledge in the minds of experts, is rarely quantifiable, and is accounted for qualitatively and subjectively during an expert's interpretation of the underlying data. Our definition of implicit error fits into the broader taxonomy of uncertainty by Boukhelifa *et al.* [147], contributing additional details and considerations surrounding their notion of data uncertainty.

Implicit error has two core components that are important for interpretation and analysis. The first is a set of characterizing traits: the *source, type, magnitude, direction, confidence,* and *extent* of the error. These traits support downstream exploration and visual analysis, as well as computational analysis and modeling of the error. The second component is a contextualizing, semantically rich description of an expert's knowledge of the error. The contextual information is important for validation of the error as well as for sharing knowledge of the error across experts. We describe each of these types of components in turn.

5.5.1.1 Characterizing Traits

During our analysis we identified three **sources** of implicit error. The first source, *inconsistency*, describes idiosyncrasies of the data generation pipeline, or a characteristic of the pipeline that varies across pipeline implementations. In the case of the Zika epi-data pipeline, examples of inconsistencies are: "the union in area X goes on strike often and doesn't report epi-data"; "country X reports all confirmed and suspected cases as confirmed cases"; and "country X overhauled its surveillance system leading to a sudden increase in detected cases."

The second source of implicit error, *gray data*, describes reputed data that are omitted at some stage of the data generation pipeline, due to, for example, standardization methods. An example of gray epi-data is: "we knew that there were more cases of X in the region; however, we didn't have the infrastructure in place to include them in the report." The notion of *gray literature* is well established and highly valued within the medical community and refers broadly to findings produced and published outside of traditional academic venues [145]. The analogous gray data are gaining traction within global health as unofficial surveillance and reporting mechanisms, such as citizens reporting on cases via cell phones, are increasingly seen as effective, rapid early alert and predicting systems [148], [149].

The third source of implicit error, *retrospective adjustment*, describes downstream updates to previously reported data, resulting in temporal discrepancies. As described in Section 5.4.1, epi-data are published weekly as static reports and thus, as a consequence, updates and modifications can only be implemented downstream. The regional level WHO data address this issues by publishing footnotes highlighting these errors with different levels of contextualization. Examples include: "after retrospective review, laboratory-confirmed cases were adjusted by X's Ministry of Health as of 25 August 2016" and "X number of confirmed cases were reclassified as suspected." These footnotes help explain questionable trends in the data, such as a sudden drop in cumulative confirmed cases, but also qualify otherwise reasonable and potentially important events, such as a spike in suspected cases.

Characterizing the source of an error is often critical to correctly interpreting the error **type**; this type can be either systematic or random. For example, "unreported data due to a strike by union workers" is likely a random error, whereas "reported confirmed cases are delayed due to lab capacity" is likely systematic. Identifying the type is important as systematic errors can often be reduced in downstream modeling or through adjustments to the data generation pipeline itself [150].

Implicit errors can also be characterized by their **direction** and **magnitude**. Direction describes the *sign* of the difference between the reported value and the value adjusted to account for the error, whereas magnitude describes the *size* of this difference. Magnitude characterizations can be quantitative, but in the case of epi-data they are most often qualitative. An example of this is "reported confirmed cases really just shows the tip of the iceberg." Furthermore, the implicit error may also have an associated measure of **confidence**, which describes domain experts' confidence in their knowledge of, or their degree of understanding about, the error. In global health, this measure of confidence is often qualitative, such as "I have a hunch that this is happening, but I don't have all the details." The direction, magnitude, and confidence of an implicit error support downstream models for analysis, regardless of whether such models are computational or mental.

Finally, the **extent** of an implicit error describes the data that are impacted. An error can impact a single measurement, a set of associated measurements, or all measurements. In the case of epi-data, the extent relates to which indicators, over what geographic area, and during what temporal window. An implicit error could, for example, reflect on a single reported case measurement, all case measurements associated with a specific indicator, or all case measurements reported for a geographic area.

5.5.1.2 Contextualizing Descriptions

Although the traits of an implicit error are valuable for visual analysis and modeling, they lack the rich contextual description that is important both for validating the trustworthiness of the error and for transferring an expert's domain knowledge to other analysts. For example, the traits of an implicit error could be characterized as follows:

- *source*: inconsistency
- *type*: systematic
- *direction*: negative
- *magnitude*: unknown
- *confidence*: very certain
- *indicator extent*: number of cases of Zika in pregnant women
- *geographic extent*: country X
- *temporal extent*: all weekly reports

This characterization, while useful for analysis and modeling, lacks important reasoning behind the existence and knowledge of the error.

More insightful is a description that includes expert knowledge that contextualizes the specific error: "Country X only reports cases of Zika in pregnant women detected within the first trimester." This description provides specific insight into the nature of the error and context for reasoning about why the error exists and the impact that it has on the reported values. In cases where domain experts are misinformed or biased, or in cases of conflicting knowledge across experts, descriptions such as this, along with stated measures of confidence as described in Section 5.5.1.1, will enable experts to evaluate reported errors against their own contextual knowledge in order to assess credibility, reliability, and impact [147].

5.5.2 Externalizing and Analyzing Implicit Error

Externalizing expert knowledge about implicit error and its surrounding context is an important first step toward understanding the nature of implicit error within a given domain, differentiating between systematic and random errors, developing models that account for systematic errors, and designing appropriate mitigation strategies for the data generation pipelines themselves. Whereas the externalization of traits of implicit error can support interpretation by visualizing the traits alongside the data, easing the cognitive load of an expert analyst, the externalization of contextual descriptions assists in validating and synchronizing expert interpretations.

For the purposes of this work, we define **externalization** as the *capture*, *characterization*, and *contextualization* of implicit error. We use the term **capture** to describe the indication of an implicit error by a domain expert. Once an implicit error has been captured, an expert **characterizes** the error by specifying, to the highest precision possible, its traits: source, type, direction, magnitude, confidence, and extent. The expert **contextualizes** the implicit error by explicating the relevant, contextual information about the source and nature of the error, as well as how it should be interpreted alongside the data.

The footnotes included in the WHO regional and country level epi-data exemplify initial efforts within global health to externalize implicit error and report it alongside official data. Similar examples of footnotes are found in other established global health datasets. These footnotes largely capture the characterizing traits of the error but usually do not include much, if any, contextualizing description. An example of a published footnote is: "As of 29 December 2016, the number of suspected cases decreased based on the modification by the Ministry of Health for Country X." These footnotes, which capture only a small percentage of the known implicit errors in epi-data, served as initial inspiration in the research reported in this work as they both acknowledged the presence of implicit error and inspired annotation as an effective externalization and visual analysis mechanism.

Building on the footnote idea, we sought to develop a structured process for enabling experts to externalize implicit error in a general and descriptive way, which we could then codify in a tool. For guidance, we turned to existing epistemological frameworks from information sciences and knowledge management [142], [151], as well as adaptations and extensions of these models developed within the visualization community [126]–[128].

The DIKW pyramid defines the relationship between data (D), information (I), knowledge

(K), and wisdom (W) [143], [152], [153]. *Data* consist of measurements that have no particular meaning in and of themselves. Contextualized data form *information* that conveys meaning. When combined with personal perceptions and previous experiences, information is transformed into *knowledge*, which supports evaluating and incorporating new experiences and information. The transformation of knowledge into *wisdom* is marked by the ability to identify and analyze patterns in one's knowledge base in order to extrapolate and make predictions. The DIKW pyramid can be also inverted: knowledge can be externalized and transferred between people as information, and information can be captured and stored as data. This inverted view of the DIKW pyramid maps to our goals of externalizing experts' knowledge about implicit error into both contextualizing information and data traits.

The DIKW pyramid provides insight into the formal relationship between knowledge, information, and data, but work within the visual analytics community models *how* knowledge can be externalized and analyzed using a visualization system — referred to as knowledge-assisted visual analytics. These models argue for the effective role of visualization to facilitate insight by illustrating how expert knowledge interacts with data through a mediating visual representation [43], [126]. This interaction is key for externalizing knowledge of implicit error, as well as for incorporating it into the visual analysis pipeline.

These models define how concepts of data, information, and knowledge in perceptual-cognitive space can be translated to, and represented in, computational space [127], [128]. The models describe a process for externalizing knowledge [126], [128] that incorporates mechanisms for both direct externalization, such as through an annotation interface, and indirect externalization, such as through interaction mining [126]. These externalization process models, however, omit the concept of information, which plays an important, contextualizing role in the externalization and analysis of implicit error. Based on these models, we derive a process model for implicit error that incorporates information.

The process model is presented in Fig. 5.2. As in previous models, circles denote processes, squares denote storage containers, and the model is divided into computational and perceptual-cognitive spaces. The model describes three stages: *identify* the existence of an implicit error through the use of a visualization system; *externalize* the implicit error through an annotation interface; and *analyze* the errors through incorporation into the visualization system. More specifically, the **identify** stage resembles the traditional interactive visualization process: data $\boxed{D}$ are

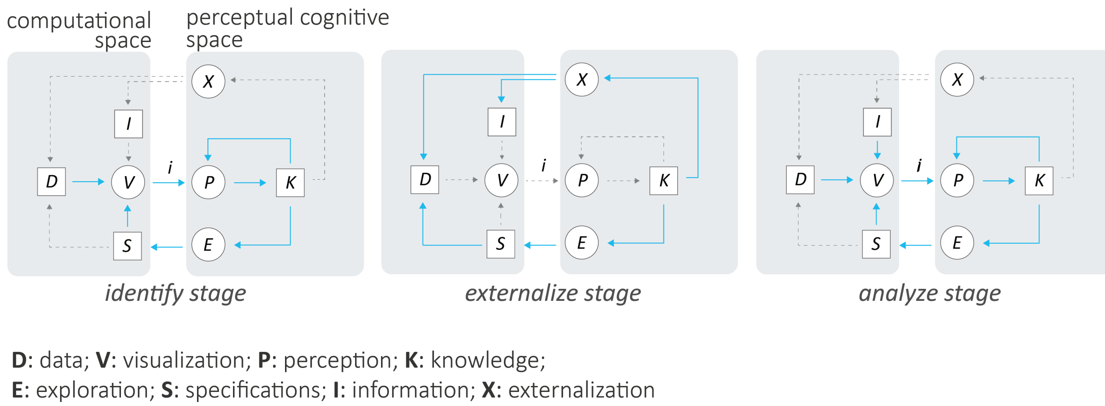

Fig. 5.2. Process model for externalizing implicit error. The model operates in three stages: In the **identify** stage, insight about the existence of implicit error is generated through the use of a visualization system; in the **externalize** stage, knowledge surrounding implicit error is externalized through an annotation interface; in the **analyze** stage, externalized implicit errors are incorporated into the visualization system for further analysis. This model is derived from process models for visualization and knowledge-assisted visual analytics [43], [126]–[128]

visually encoded $\textcircled{V}$ given a set of specifications $\boxed{S}$ and transformed into images i, which an analyst interactively explores and interprets through perceptual and cognitive processes $\textcircled{P}$. These processes are both informed by an analyst's knowledge $\boxed{K} \rightarrow \textcircled{P}$ and yield new knowledge in the form of insights $\textcircled{P} \rightarrow \boxed{K}$. Here, visualization provides a powerful mechanism for leveraging expert domain knowledge to generate insight about the data [43], [126]. When these insights indicate the presence of an implicit error, knowledge about the error is captured, characterized, and contextualized in the **externalize** stage via an annotation interface $\textcircled{X}$. The contextual description about the error is stored as information $\boxed{K} \rightarrow \textcircled{X} \rightarrow \boxed{I}$, and the characterizing data traits are stored as data $\boxed{K} \rightarrow \textcircled{X} \rightarrow \boxed{D}$. Some of the data traits, like the extent of the error, can be inferred indirectly from the state of the visualization system, such as where a marker is placed: $\boxed{K} \rightarrow \textcircled{E} \rightarrow \boxed{S} \rightarrow \boxed{D}$. Finally, in the **analyze** stage, the externalized error is incorporated into the visualization system for exploration and interpretation alongside the underlying data. This final stage supports the validation, synchronization, and analysis of the error by analysts. Using this process model, we designed a visualization tool for externalizing implicit error in Zika epi-data, discussed in the next section.

5.6 Instantiating the Framework

As an example of how the framework can be used in practice, we developed a prototypical system for global health response coverage assessment that supports externalizing implicit error in

Zika epi-data. The system, shown in Fig. 5.3, was built using D3 and Leaflet Javascript libraries and was designed to support the three stages of the process model presented in Section 5.5.2: an underlying visualization supports *identifying* errors; an annotation platform supports *externalizing* errors; and an overlaid implicit error visualization supports *analyzing* errors. We solicited feedback on the system from two Zika experts, and reflected on our experience to provide guidance for others seeking to instantiate the framework.

To support the **identify** stage, the core of the system is an interactive visualization interface designed to support exploration of the epi-data using a standard linked-view approach to visualizing geospatiotemporal data. The system, which was informed by the technology probes, supports exploration and comparison of outbreak and response data at three levels: the regional view, displaying regionally aggregated WHO data; the country view, displaying country level WHO data; and the subcountry view, displaying country and subcountry MoH data combined with geographic and demographic metadata. As we found with the technology probes, designing a visualization system assuming no implicit error results in a powerful mechanism for triggering and distilling insight about implicit error in a dataset. Epi-data are encoded in a choropleth and overlaid on an interactive basemap. The map is linked to a chart-view (Fig. 5.3d) displaying trends in epi-data indicators over time, as well as geographic and demographic metadata at the subcountry level. Toggling between indicators controls the data encoded in the choropleth. Sliders control the timestep shown in that map view and allow users to scroll over time. Response data are overlaid on the map either as glyphs in the regional and country view or as textured shapefiles in the subcountry view.

The system implements the **externalize** stage with an annotation platform. Our use of annotation is grounded both in the results of the annotation technology probe described in 5.4.2 and in a large body of literature surrounding the effective use of annotation for narrative and storytelling [154], [155], collaboration and communication [156]–[158], externalization of insights [159]–[161], and assessments of data quality [147]. Annotations are submitted by dropping markers on regions, countries, or subcountries, which brings up a semistructured annotation template. The template, shown in Fig. 5.3 (top), was designed based on the description of implicit error presented in Section 5.5.1, but using language that resonates with global health experts. Information about contextual descriptions of implicit error is captured and stored as unstructured text. Data about characterizing traits are selected using check boxes and radio buttons, with the exception of the *region* and *indicator* fields, which are suggested based on the current system settings. Submitted annotations are stored

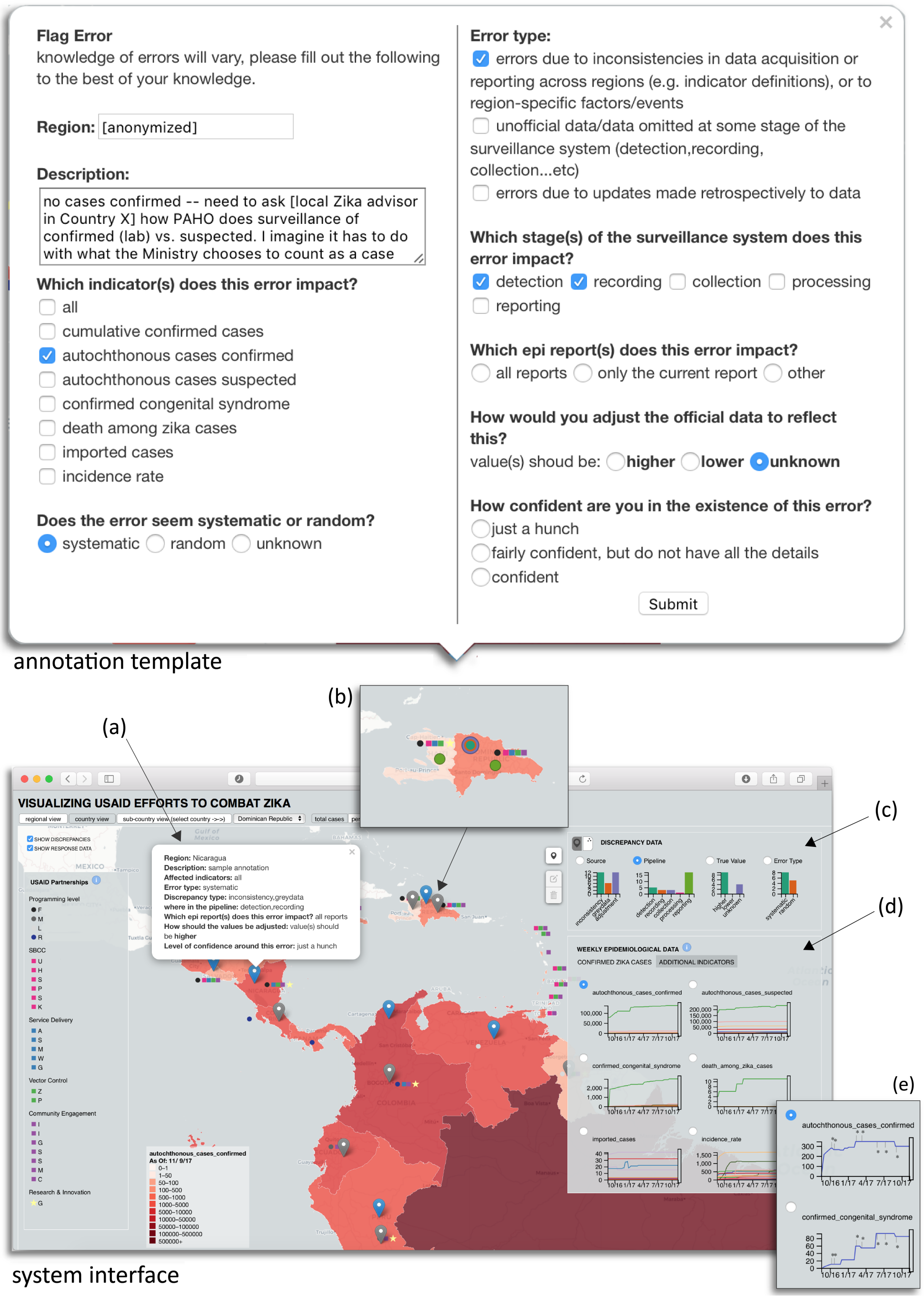

Fig. 5.3. Prototypical instantiation of the framework for externalizing implicit error. (Top) Expert knowledge surrounding implicit error is externalized via an annotation template, shown here featuring an example annotation discussed in Section 5.6. (Bottom) Once submitted, annotations are displayed either (a) as popup markers in *information mode* or (b) as bullseyes encoding categorical data trait attributes in *data mode*, which are linked (c) to a scented interactive legend displaying the distribution across categories for each trait. In information mode, annotations can additionally be viewed (d) in the chart view (e) as footnotes annotated along line charts.

in an online database. New annotations are synchronized to support remote collaboration.

Once annotations have been collected, the system supports the **analyze** stage by visualizing the submitted annotations using established encoding techniques. Here, visualization supports the identification of patterns, outliers, and correlations within the externalized error and in relation to the original dataset. More specifically, the information and data stored in the annotations are presented via two different modes. *Information mode* displays annotations in the form they are submitted — as popup markers displaying data traits and contextual information (Fig. 5.3a). Markers are filtered by view, such that regional level annotations, for example, appear only in the regional view. Contextual information is additionally annotated along the line charts in the chart view (Fig. 5.3e), a feature that proved helpful in the technology probes for verifying potentially significant spikes and other trends in the data. In *data mode*, annotations are instead encoded as circles, color encodes categorical attribute corresponding to a single trait, and traits belonging to multiple categories are encoded as bullseyes (Fig. 5.3b). Users can toggle between traits through an interactive legend [162], scented with the distribution of categories for each trait [163] (Fig. 5.3c).

The system is designed for long-term individual and collaborative use by Zika experts. We received feedback from two experts who used the tool collaboratively in a guided, think-aloud interview: one is a coauthor of the publication of this design study. The feedback indicated that in the short term, the system provides a platform for discussing, reasoning about, and formalizing an understanding of implicit error. For example, after submitting an annotation, one of the experts commented that interacting with the visualization made her think about the data: in cases where she initially questioned the data, it compelled her to reason about why the data were in fact correct, or alternatively, what kind of error could account for what she was seeing. The feedback also suggested that longer term, the system would be valuable for developing a database of externalized implicit errors — or, as one expert put it, an "institutional memory" — which could provide a platform for modeling error and informing mitigation strategies.

In reflecting on our experience of developing this system, we offer one possible strategy for others looking to instantiate the implicit error framework:

- Start by designing a visualization system assuming no implicit error. We found that designing for the ideal data scenario, which involved building a problem characterization and data and task abstraction, identifying design requirements, and prototyping design solutions assuming the perfect dataset, provided a valuable platform for investigating implicit error within the data. Engaging domain experts in discussions surrounding the perfect dataset facilitated discussion on discrepancies within the existing data generation pipelines, as well as how these discrepancies impacted their analysis and their general attitude toward data and visualization. We imagine there could be cases in which visualization researchers and domain experts want to incorporate the IE framework into existing systems, e.g., as follow-ups to previously completed design studies. In such cases, we encourage revisiting initial problem characterizations and abstractions through the lens of implicit error.

- Use the results of the first phase, described above, as technology probes. We found that the prototypes developed in the initial stage of the design study provided valuable technology probes for deeper investigation: for gathering examples of implicit error, for investigating how domain experts incorporated implicit error into their analysis, and for identifying potential mechanisms for externalization. We employed these probes both informally in early discussions with our primary collaborators, and formally with larger groups of domain experts. The results of the probes informed our characterization of implicit error and the design of our annotation mechanism, from the interaction design to translating the framework into a language that resonated with domain experts.
- Approach the visualization of implicit error as a separate design study. In the case of Zika outbreak data, and, as we speculate, extends more broadly, the problem characterization, data and task abstraction, and design requirements for visualizing implicit error are often perpendicular to those of the underlying Zika outbreak data and requires separate consideration. In addition, as implicit error accumulates through the use of an externalization mechanism, understanding of the problem, data and tasks, and design requirements will evolve and change, and the visualization design should evolve and change accordingly, e.g., to support emergent tasks or scalability issues that arise.

5.7 Discussion

The framework for externalizing implicit error proposed in this chapter was inspired by, and grounded in, a design study with global health experts studying Zika epidemiological surveillance data. Implicit error, however, is prevalent in a broad range of fields. We cite discussions surrounding implicit error from a number of different fields in Section 5.2. We further speculate that other domains that rely on distributed, heterogeneous data generation pipelines also feature implicit error.

For example, bioinformatics increasingly relies on, and requires that, independent teams of researchers publish datasets alongside academic papers so that others can replicate and build on the results. Nuances of the data generation pipeline for individual datasets may be critical for ensuring the reliability of results. Another example is monitoring air quality conditions around the globe, where sensor networks deployed by individuals, grassroots organizations, academics, and government agencies largely function independently. Improving the scientific understanding of air quality, as well as impacting policy changes, requires integration of these networks. However, differences in the types of sensors and how they are deployed, reliability of the data collection system, and local environmental and cultural conditions make meaningful standardization of the data a challenge without documented knowledge of these variations and their impacts. These are just two examples of potentially many that could benefit from thoughtful consideration of implicit error and deployment of mechanisms to support externalization.

We argue in this work that externalization of implicit error could lead to models of systematic error that augment data, as well as refinements to the data generation pipelines themselves. Our work here is a first step toward this goal: understanding the nature of implicit error in a dataset is necessary before it can be accounted for in a robust way. We anticipate that as descriptions of implicit error accumulate, we can begin to model the error and then use the models to inform error mitigation, and perhaps even to guide institutional change in distributed data generation pipelines. It is also possible that as our understanding of implicit error evolves, we could model the cognitive-perceptual adjustments that domain experts make when incorporating knowledge of implicit error into their interpretations, and simulate these adjustments through modifications to the data. Additionally, models of implicit error could improve the reliability of manual adjustments that experts commonly make [147].

The process model for externalizing and analyzing implicit error described in Section 5.5.2 modifies existing knowledge-assisted visual analytics models [43], [126], in part by explicitly incorporating the concept of information, denoted as $\boxed{I}$ in the model. We view information as existing in between computational and perceptual space and playing an essential role in the transfer of knowledge across experts. Our position is that not all expert knowledge can be digitized completely, which is a possible limitation of existing knowledge-assisted models that rely on computable explicit knowledge. In the context of building institutional knowledge and synchronizing interpretations across experts, information is likely more valuable and powerful than a reduced computable data representation. Thus, we believe that the addition of information into these models could be a useful perspective for designing future knowledge-assisted visual analytics tools.

In reflecting on this work, we argue for the value of field studies in uncovering new, and unexpected, visualization opportunities. Taking full advantage of a field study, however, requires careful and thoughtful navigation of interpersonal and interorganizational relationships. We offer several recommendations based on our experiences reported in this chapter. First, when conducting field studies in large, organizational settings, we encourage researchers to take advantage of the larger community and to interact with a range of stakeholders in order to identify challenges that extend more broadly. Second, we recommend presenting talks on applied visualization research early on, as this will help shape a visualization researcher's role and expectations of the collaboration. Third, organizations may expect deliverables that do not neatly align with the nature of design study — we found that developing a visualization tool early-on that was valuable to the organization helped to establish credibility and to obtain the buy-in necessary to pursue design study research.

Finally, we used this project as an opportunity to investigate new ways of approaching, recording, and reporting design study research. By approaching the development of visualization prototypes as an opportunity to probe and learn, as well as committing to extensively recording insights and artifacts throughout the project, we found that design study offered us new ways to discover and reflect on the visualization needs of domain experts. We felt it necessary to report our process through a rich description in order to capture the value that design study nuances have for the knowledge we acquired, but we note that we were careful to cull our descriptions to just those details we felt had an impact on our findings. We hope this work adds to the conversation on the role of design study in developing and refining visualization knowledge.

5.8 Reflection

In the following section, we reflect on our experience conducting this design study through the lens of ADR. We highlight the ways in which ADR strengthened our research, along with our efforts to extend the framework to better suit applied visualization research. We additionally discuss several strategies that proved helpful in supporting ADR throughout the design process.

5.8.1 Strengthening the Research Process with ADR

Approaching this design study through the lens of ADR primed us to pay close attention to research context, participants, and the broader domain expert community. Understanding and attending to these elements proved critical to the success of the project. We entered the field study with very little knowledge about who we would be working with, the kinds of problems we would encounter, or what data even looked like in the global health context. Being completely embedded within the domain offered a rich research environment for identifying interesting domain problems and visualization opportunities, and for selecting promising research collaborations [8]. In addition, being surrounded by domain experts working in different areas helped us to understand how specific problems extended more broadly, and to identify general domain themes and challenges. Whether consciously or not, the insights that we gained through discussions and interactions with the broader organization, specifically surrounding the general relationship and attitude toward data, heavily shaped our research throughout the study.

Conducting research from within the domain also introduced a certain amount of overhead and required an additional degree of navigation and diplomacy. Global health work is high-stress, fast-paced, hierarchical, dynamic, and deliverable-oriented. Priorities, teams, and expectations often change with changes in funding and/or current events. Design study research, on the other hand, is slow and requires a substantial commitment from domain collaborators. Additionally, this research is emergent and insight-driven, and deliverables cannot be stated up front or even guaranteed. Meeting the organization's expectations of the field study, obtaining buy-in from gatekeepers [8], and integrating our research into the organizational work flow required careful navigation and a heightened sensitivity to organizational and interpersonal dynamics. ADR does not offer explicit guidance on navigating the intricacies of large organizational and corporate settings, but the framework's general emphasis on the role and impact of people and context left us better prepared to conduct this research.

Throughout the design process, ADR provided a useful structure for our learning, framed around cycles of intervention interspersed with critical reflection. In this design study, we viewed technological intervention primarily as a mechanism for learning about the domain problem. We were less focused on iteratively designing and developing a deployable technology artifact, although we suspect that given a longer term collaboration, the focus of our cycles would have eventually shifted. We conducted two and a half cycles of intervention. Each cycle was framed around the learning that we wanted to accomplish, and we designed technology probes to facilitate this learning. The first cycle was devoted to investigating how our collaborators interacted with and interpreted their data. The second cycle was devoted to investigating expert knowledge of discrepancies in the data, and the ways in which visualization helps to externalize and incorporate this knowledge. The final half-cycle, which followed a period of formalization of learning, was devoted to testing the expressivity of our framework and to starting to investigate effective approaches to visualizing implicit error. We viewed our final technology probe as a proof of concept of our framework; however, it also serves as a snapshot of our learning and a starting point for future cycles.

Our cycles of intervention were guided by ADR's principles of reciprocal shaping and mutually influential roles. Whereas reciprocal shaping emphasizes the inherent and valuable role of context and participants in shaping the research process, mutually influential roles promote the learning that should occur on both sides of the collaboration. Much of the reciprocal shaping was subtle and persistent. There was a constant negotiation between our collaborators' ambivalence toward using data and visualization for analysis and our desire to be data and visualization evangelists to show our collaborators how valuable and revealing data visualization can be. More explicit reciprocal shaping occurred when a new primary collaborator with interests in data discrepancy prompted us to pursue the topic more fully.

In adherence to mutually influential roles, disruption of thinking on both sides of the collaboration became a driving force of our research. On the domain side, the ultimate goal of the design study was to shift the domain experts' relationship with data and visualization – from viewing data as a semiaccurate reporting mechanism that could be misconstrued or even weaponized, to viewing data as empowering and viewing visualization as a powerful mechanism for insight and meaning-making. On the visualization research side, this work pushed us to re-evaluate the role and value of visualization within the given context, causing us to shift our view of visualization from a tool for meaning-making to a mechanism for triggering expert knowledge. In addition to disruption within the collaboration, our notion of implicit error seeks to disrupt current thinking about error and uncertainty within the visualization community.

Adhering to the principle of mutually influential roles strengthened our research in other ways as well. It encouraged us to pay more attention to the backgrounds and perspectives of our collaborators, leading us to, among other things, the fundamental insight that our collaborators' domain knowledge included an in-depth understanding of data discrepancies. Adherence to mutually influential roles also encouraged us to pay attention to and to leverage the individual roles of our collaborators and the different ways in which each collaborator could contribute to the research.

Following our first two cycles of intervention, ADR provided structure for formalizing a broader span of learning, and gave us the confidence to promote this learning as the primary contribution of our work. Unlike with the Poemage design study, in which formalization of learning occurred as an afterthought while preparing the publication, a significant portion of this design study was devoted to formalizing our learning about implicit error into an abstract, generalizable framework. We view our framework as an example of the ways in which learning from reciprocal shaping and mutually influential roles can be formalized and generalized.

5.8.2 Extending ADR for Applied Visualization Research

In our preliminary investigation of ADR, we questioned the generalizability of outcomes stemming from deliberately subjective and emergent research environments. We argued that researchers could, at best, only speculate about how their findings might generalize. Following the design study methodology [8], we proposed the notion of *transferable outcomes* as a potentially more suitable alternative to the generalized outcomes emphasized in ADR. Whereas generalized outcomes cast specific findings as instances of broader classes of findings [5], transferable outcomes provide sufficient contextual detail such that others can adapt and transfer specific findings to their own research contexts [21].

After further consideration, we posit that such generalizations, while primarily speculative, still hold value for the visualization community, and that applied visualization researchers who are deeply embedded within a research context are still the best suited to speculate. We see transferability instead as a way to increase the value and impact of applied visualization research. We therefore used this design study as an opportunity to explore ways of increasing transferability of applied visualization research. As part of our publication, we included a rich and reflective description of our research context, our process, the development of artifacts, and our learning. Our description was based heavily on field notes documenting our reflection and learning throughout the design process. Summaries of these field notes were also made available in an interactive timeline included as supplemental material. Our hope was that this added contextual detail will help in two ways. First, it would increase the reliability of our research by allowing readers to more thoroughly scrutinize and verify our process and findings. Second, it would aid our fellow researchers in identifying similarities and differences between our work and theirs, and in adapting potentially useful elements of our research accordingly.

5.8.3 Supporting ADR Throughout the Design Process

Throughout this design study, certain strategies emerged as particularly helpful for supporting ADR in an applied visualization research context. First, we found that giving our collaborators agency to promote their perspectives helped to create a safe and conducive environment for reciprocal shaping. It was important for our collaborators to feel that their views and ideas, no matter how radical, were valued and taken seriously. As we discovered in our Poemage design study, this agency can increase the trust and engagement of collaborators, and can also lead to visualization insights and opportunities. In addition to granting this agency, being sensitive and attentive to moments of subtle reciprocal shaping allowed us to identify and act upon important themes that we otherwise could have missed. In both design studies, we found that the casual experimental nature of technology probes helped to promote this agency. In this design study, technology probes also proved valuable for facilitating the different cycles of intervention and learning.

Second, we found that field notes were a valuable mechanism for reflection and learning. In this design study, we embraced the notion that applied visualization research is insight-driven. Insights emerge through interactions, design and development, literature reviews, and reflection and guide visualization researchers along existing and new lines of inquiry. Seeking guidance from field research methods [164], we adopted a practice of taking field notes to facilitate our reflection and learning. We structured our field notes around insights and human interactions, no matter how small or seemingly insignificant. In each of these cases, we tried to capture and reflect on the role and influence of context and participants. We additionally tried to note any *relevant characteristics* of the context of participants, and to reflect on how these characteristics were unique or overlapping with other research contexts. We also used field notes to store our reflective transcriptions of recorded meetings and interviews. Reflective transcription was a technique that had proven valuable for promoting reflection and insight in past collaborations, and which we formalized in this design study. Field notes helped to focus and systematize what initially felt like an unconstrained and overwhelming task. Field notes had the added benefit of documenting our reflection and learning, providing ample raw material to support formalization of learning and providing the rich trail of evidence that we used in our efforts to increase transferability and reliability of our research.

CHAPTER 6

DESIGN STUDY: ANALYZING REAL AND SIMULATED GALAXY OBSERVATIONS

The fourth piece of work in this dissertation includes the initial results of an on-going design study with astronomers and astrophysicists. We viewed this design study as an opportunity to further test and validate our results of applying and extending ADR in our design study with global health experts, presented in Chapter 5. Once again, we paid close attention to the dynamics and perspectives of all members of the research team, domain experts and visualization researchers, and allowed moments of hesitation, inclination, and disruption to guide the research process. We framed our learning around cycles of intervention interspersed with critical reflection and employed an evolving technology probe as a central mechanism for these cycles. We continued our practice of taking field notes in order to support reflection and learning throughout the design process and to increase transferability and reliability of our research.

As part of a continued effort to increase transferability and reliability of research, we included a rich process description as supplemental material alongside the publication of our early-stage results. Whereas early-stage research focused on the design of a visualization platform, we began to pursue broader, emergent lines of inquiry surrounding mismatches between visualization guidelines and domain practices. We formalize the learning surrounding these lines of inquiry as important areas for future visualization research.

We additionally used this design study as an opportunity to experiment with new ways of supporting ADR throughout the design process. Drawing from field research methods, we extended our practice of taking field notes by being more purposeful about including and differentiating between *descriptive* information — documenting factual data about context and interactions — and *reflective* information — documenting questions, interpretations, and ideas [164]. In addition, we developed a second, *visualization research* technology probe to experiment with supporting field note taking throughout the study.

In what follows, we present the details of our early-stage research. We then reflect on our experience of further testing and validating ADR within a visualization design research context, and on the results of our experiments.

6.1 Overview

When we look out in space, we are looking back in time. Astronomical observations of galaxies, like those taken by the Hubble Space Telescope, provide snapshots at different stages of galaxy formation and evolution, as if billions of flip books were torn apart and spread across the sky. Theoretical astrophysicists develop models that effectively reassemble these flip books. Through a sequence of simulated snapshots, these models predict how galaxies form, interact, merge, and evolve over time. Matching *real* observations from telescopes with *simulated* observations from snapshots of theoretical models can help astronomers predict the histories and futures of individual galaxies. Comparing collections of real and simulated observations can help validate models and clarify key processes in galaxy formation and evolution.

Astronomers and astrophysicists studying galaxy formation and evolution analyze and compare these data through a combination of statistical analysis of plots and visual analysis of images and movies. These two forms of analysis are most often completed independently or sequentially. Integrated analysis of these data is not supported in existing astronomical data analysis tools. The goal of this work is to begin to fill this gap by exploring integrated approaches to statistical and visual analysis to enhance the analysis of these data.

This work evolved out of a long-term, collaborative relationship with astronomers and astrophysicists at the University of California Santa Cruz. We report here on the initial results of an ongoing design study. Current contributions of the study include a data and task abstraction for statistically and visually analyzing real and simulated galaxy observations and an initial design, implemented in a prototype called GalStamps and evaluated through two case studies with domain experts. Our results lay the groundwork and introduce interesting visualization research opportunities for further design study research. As supplemental material, we include a rich process description in order to support transparency and transferability of our research.

6.2 Problem Domain Background

One of the major questions in astronomy is how galaxies were formed and how they have evolved over time. Observational astronomers study galaxy formation and evolution by analyzing observations of galaxies captured by space-based and ground-based telescopes. These telescopes survey regions of the sky, imaging galaxies with instruments that measure different wavelengths of light. Astronomers then derive additional statistical measurements from these images to characterize and analyze the observations.

Theoretical astrophysicists, on the other hand, study galaxy formation and evolution by developing computational models that simulate how galaxies form, merge, and evolve over time. Comparing the results of these models to real observations helps researchers validate and improve models and clarify what is being exhibited in observations. To support this comparison, theoretical astrophysicists generate collections of *simulated* observations by taking snapshots of their simulations at different angles and different stages of evolution, rendering images at different wavelengths of light, and degrading the images to resemble those taken by telescopes. Similar to real observations, comparable statistical measurements are then derived from these simulated images. Astronomers have termed these real and simulated images *galaxy postage stamps*.

In both cases, scientists analyze galaxy postage stamps and derived statistical measurements in order to identify trends, correlations, and outliers that help to answer open questions about galaxy formation and evolution. Analyzing statistical measurements reveals features that are either not captured or are less evident in the images, and vice versa. Visually inspecting the images associated with features in the statistical measurements allows scientists to verify their findings and to discover correlations between statistical and visual attributes. Finally, comparing features across real and simulated collections of observations is critical to these scientists' analyses: agreement in features allows scientists to validate existing models, and thus the underlying theory, whereas disagreement in features helps scientists refine theory and even uncover new, key phenomena. Neither of these forms of analysis, visually verifying statistical features and comparing features across real and simulated collections of observations, is supported in an integrated way by existing astronomical pipelines and technologies. As a result, these powerful techniques are used only on occasion, which we speculate has resulted in missed opportunities for verification and discovery.

6.3 Related Work

Within the astronomy and astrophysics community, a host of services have been developed to support the dissemination and analysis of real and simulated observation data. These platforms support the navigation and exploration of astronomical surveys and catalogues [165]–[168] and offer a range of tools for mapping and cross-matching observations across datasets [169], [170], statistically analyzing associated tabular data [171], and visually classifying real and simulated galaxy observations [172]. Numerous visualization packages have also been developed to support statistical and visual analysis of astronomical data more generally [173]–[175]. These packages target a broad range of data types, from catalogue data to time-evolving numerical simulation data, and support an equally broad range of visualization techniques, from volume rendering to 2D and 3D scatterplots. Although some of these tools support both statistical and visual analysis, these tasks are typically performed separately.

Within the visualization community, research has mainly focused on the *visual* representation of astronomical data. This work spans efforts to visualize astronomical and cosmological simulation data [176]–[178] to reconstructing three-dimensional (3D) models from two-dimensional (2D) observations [179]–[181], to developing 3D Universe software, which visualizes and maps observation data from a range of sources for use in planetariums and other public outreach venues [182], [183]. Other instances of visualization research focus on representing uncertainty in astronomical data [184] and supporting the mapping and cross-matching of multisurvey observational datasets [185]. These instances focus purely on visual analysis and not on statistical analysis.

A handful of successful approaches have combined visual and statistical analysis of astronomical data using linked views [186]–[188]. This work, however, falls primarily in the context of visualizing and analyzing cosmology simulation data. A more general push for linked view analysis of high-dimensional astronomical data has come from the astronomy visualization community [189]. Glue [190], a widely used Python library designed to support multidimensional linked-data exploration, is a direct response to this call. With Glue, users can manually generate scatterplots, histograms, and images from multiple related datasets. The platform is built around brushing and linking, such that selections in any one graph or image are propagated to all others. Glue, however, focuses its analysis on individual images, allowing users to query and visually inspect image pixels corresponding to statistical properties. Our work extends the Glue framework by supporting this analysis across collections of images.

6.4 Process, Artifacts, Reflection

As part of an ongoing investigation into recording and reporting on design study research, we include a rich process description detailing our methods and resulting artifacts, and reflecting on the learning that occurred along the way. We note that the following content was originally included as supplemental material alongside the publication of our work, with a high-level process summary (omitted) included in the main text of the publication.

As shown in Fig. 6.1, the design study was carried out in three core phases (P1-P3) structured around five visits (V1-V5) to the University of California Santa Cruz (UCSC), concluding with a period of remote, iterative design and development. Our collaborators consisted of faculty, graduate and undergraduate students, and a postundergraduate researcher, all conducting research in either observational astronomy or theoretical astrophysics and with a focus on galaxy formation and evolution.

6.4.1 Preconditioning and Problem Formulation

Preconditioning and problem formulation [13] occurred over the course of two visits separated roughly 18 months apart in the 2 years prior to the start of the design study. Research between visits was minimal, and the visits were conducted in parallel with other visualization research projects. Therefore, we do not consider this a core phase of our design study.

The primary visualization researcher was already deeply familiar with the domain and had an established relationship with domain expert collaborators, having studied physics at the University of California Santa Cruz (UCSC) and subsequently worked for the collaborators in computational astrophysics and astronomy visualization. We were motivated to conduct a design study in the domain as it offered a rich space to explore approaches to combining spatial and abstract visualization techniques. We entered this initial period of the design study with a general sense of problem space. Our goals during these two visits were threefold: 1) to scope out a project that met the criteria for a successful design study [13]; 2) to understand the broader research context, including the roles and

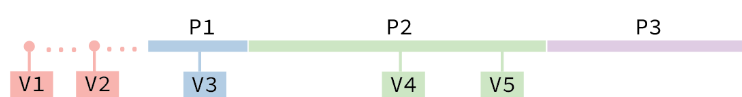

Fig. 6.1. Phases of the research process. The design study was conducted in three core phases (P1-P3) and structured around five visits (V1-V5) to the domain experts' home institution.

dynamics of the group, how research was conducted, and our collaborators' current relationship with data and visualization; and 3) to clarify our role as applied visualization researchers, the objectives of our research, and the design study process.

During the visits, we conducted unstructured and semistructured interviews with domain faculty, graduate students, and undergraduate students. We met with five faculty members, seven graduate students, and three undergraduate students. Interviews were conducted individually and in groups and were nearly always recorded. Interviews were structured around learning about current research projects, how visualization was being used to answer research questions, and the questions that domain experts were struggling to answer using existing approaches. During these visits, we also attended and participated in group meetings, attended talks and colloquia, and read relevant literature.

6.4.1.1 Artifact

The primary artifact of this phase was the problem formulation. The problem initially emerged in a group meeting with students and faculty during the second visit. A student was giving an update on his research. The group was examining a scatterplot, singling out a subset of points corresponding to a potentially interesting feature in the data. A faculty member said something along the lines of "we should take a look at the corresponding galaxy images." Under the group's current analysis pipeline, looking at these images would have required a lengthy process of identifying the points of interest, pulling the associated FITS files [191] (likely from a remote cluster or supercomputer), and running a Python script to render the appropriate images. Allowing domain experts to interactively query galaxy observation images via selections in plots would expedite this process and could greatly enhance their analysis.

The group meeting occurred at the end of the first day of a 3-day visit. The remainder of the visit was spent triangulating and further investigating the domain problem via discussions, brainstorming, and sketching with faculty, graduate students, and undergraduate students. We define the domain problem at three different levels: *high-level:* help understand how galaxies were formed and how they have evolved over time; *mid-level:* help verify computational models and clarify what is being exhibited in observations; and *low-level:* support interactive analysis and comparison of real and simulated galaxy observations. Querying images via selections in associated metadata forms one of the core tasks of this interactive analysis.

6.4.1.2 Reflection

Entering the study, our collaborators had many ideas for projects that could benefit from our visualization expertise, and they anticipated that we would be able to contribute broadly over a short period of time. It was therefore necessary that we spend some time up front discussing the components of design study research: the dual-goal of solving real-world problems *and* generating new visualization knowledge; the stages of the design process and the time and effort required by all team members; and the exploratory nature and expectations surrounding final outcomes. Because of our existing relationship with our collaborators, we were able to have these conversations more honestly and transparently than was our experience in previous design studies.

It was also necessary to extend our collaborators' view of, and approach to, data visualization for analysis. Astronomy visualization has, in large part, been dedicated to communicating scientific research and engaging the general public. This focus on presentation appears to have set the tone for what *visualization* translates to in research as well for our collaborators. Although a project centered around scatterplots and images may have seemed initially underwhelming, our collaborators quickly saw the potential value in this kind of visualization for their analysis. In addition, our collaborators tended toward a technique-driven approach to visualization: generating visualizations using the latest, state-of-the-art techniques and exploring the results for potentially new insight. Thus, our goal during these visits was also to transition their thinking to a more task- and question-driven approach to visualization design.

6.4.2 Learning About Galaxy Observation Analysis

Following the preconditioning and problem formulation, the first core phase of the design study was devoted to learning about the analysis of galaxy observation data. Our goals for this phase were to understand the data and tasks associated with analyzing and comparing real and simulated galaxy observations and to obtain an initial dataset. The phase spanned 3 months and was structured around a single visit to UCSC.

The visit consisted of more unstructured and semistructured interviews with four faculty, four graduate students, and two undergraduate students, all of whom we had worked with in previous visits. Interviews were structured around understanding our collaborators' current statistical and visual analysis pipelines. In several cases, we conducted think-alouds in which we asked our collaborators to walk us through their analysis of existing plots and images. We additionally participated in a group meeting with faculty and students in which we demonstrated existing, related tools (namely yt [175], [192] and Glue [190]), and then brainstormed, via sketching ideas on a whiteboard, how such tools could be extended to support the existing domain problem. The brainstorm session included a discussion of data and tasks.

6.4.2.1 Artifacts

The primary artifact of this phase was an initial data and task abstraction, which we continued to develop and refine over the course of the study. The final abstraction is presented in Section 6.5. This initial abstraction formed the basis for our visualization design work. A secondary artifact of this phase was an initial dataset: a collection of simulated galaxy observations. The dataset comprised a collection of images, a set of two images for each observation, and a table of quantitative data attributes derived from analyzing the images. The images were generated by taking snapshots from a set of high-resolution, hydrodynamic, cosmological galaxy simulations known as the VELA simulation suite [193]–[195]. Snapshots were postprocessed using the Monte-Carlo Radiation Transfer code Sunrise [196] to incorporate the effects of light and dust, and degraded using different noise models to reflect observations taken by telescopes. The results were then analyzed using an algorithm called GALFIT [197] in order to derive a set of quantitative attributes measuring various structural parameters. The data were stored in FITS files [191], the standard astronomical format for storing image and table data, and the format used primarily for storing real galaxy observation data.

6.4.2.2 Reflection

We approached the domain problem, and thus the initial data and task abstraction, as a way to probe the problem space in the hopes that it would lead to more interesting visualization research opportunities. As we progressed, however, it became evident that a standard linked-view approach could greatly enhance our collaborators' analysis. This presented us, the visualization designers, with a conflict: whether to pursue novel visualization design spaces, which may not be nearly as relevant or effective for our collaborators, or focus our efforts on developing a validated design that combines a handful of established approaches and could potentially lead to the confirmation or refinement of existing visualization design guidelines [13]. Although perhaps the more risky route from a visualization research perspective, we ultimately decided to go with the latter option.

We had assumed that obtaining an initial dataset would be straightforward, given the abundance of data and the data-sharing culture within the astronomy and astrophysics community. Unlike, for example, medical or public health data, astronomical data are hardly ever sensitive, confidential, or vulnerable to being weaponized. In addition, astronomy research is often publicly funded, with the requirement that the resulting data be made publicly available and accessible. We were therefore surprised that obtaining these data proved to be a challenge. Rather than being directed to a specific dataset, we were directed to numerous sources from which a representative dataset could be curated – a nontrivial task that required extensive domain knowledge and thus multiple rounds of input from our collaborators. We suspect that this process reflects how datasets are curated within this domain. Researchers have an enormous amount of data at their fingertips. In designing an experiment, there are often multiple, suitable datasets from various sources to curate from, which greatly expands the breadth of possible research, but presents challenges when trying to obtain a representative sample of the data with which to test and develop.

6.4.3 Exploring the Visualization Design Space

Once we had arrived at an initial data and task abstraction, and had obtained an actual dataset, the next phase was devoted to exploring the space of visualization design. Our primary goals for this phase were to elicit design requirements and develop an initial design. We also used this phase to continue to test, develop, and refine our data and tasks abstraction, and to probe for potentially more interesting visualization opportunities. The phase spanned 6 months and was structured around two visits to UCSC, approximately 2.5 months apart. During this phase we collaborated with four of the faculty members and three of the graduate students from our previous visits, and one new collaborator — a postundergraduate researcher — with whom we worked closely for the remainder of the study.

In preparation for the first visit, we developed a bare-bones technology probe, based on our initial data and task abstraction and the results of our preliminary discussions, sketching, and brainstorms, and featuring the initial dataset. The initial probe, shown in Fig. 6.2 *(top left)*, was developed in D3.js and supported only the basic functionality for generating scatterplots and parallel coordinate plots, and for querying images via plot selections (which we implemented via brushing). Over the course of the two visits, we used the technology probe to brainstorm and experiment with new features, functionalities, and interactions and to explore possible applications to a range of datasets. Through the iterative development of the technology probe, we were able to home in on

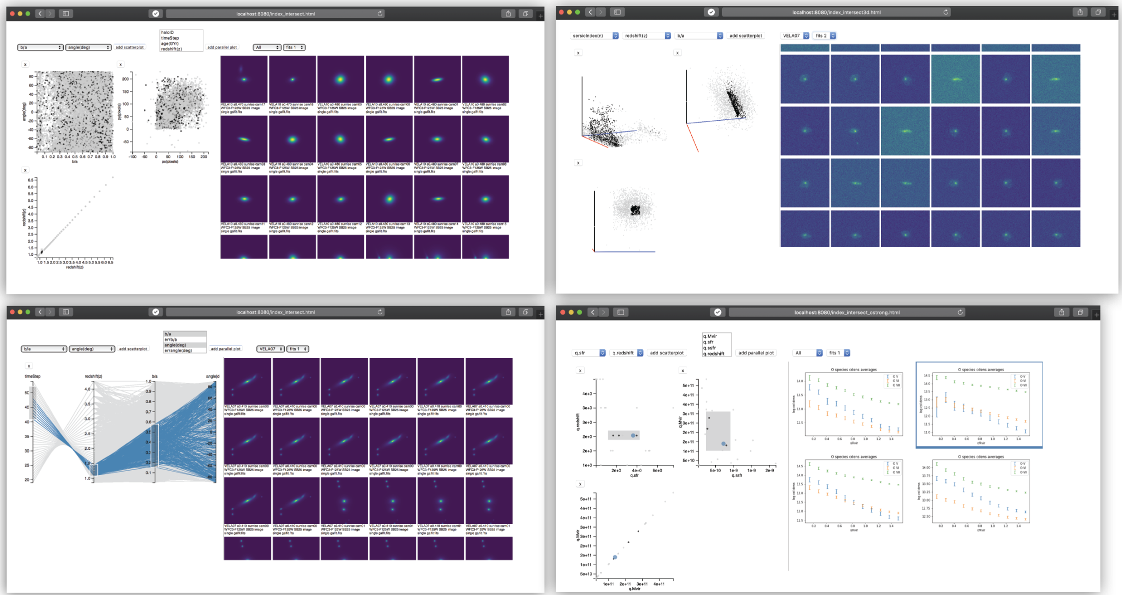

Fig. 6.2. Four iterations of the technology probe. *(Top left)* Querying images via selections in various scatterplots was a core function of the system; *(bottom left)* parallel coordinate plots proved to be effective as interactive legends for filtering data across three+ dimensions; *(top right)* interactive 3D scatterplots emerged as a potential future visualization research opportunity; *(bottom right)* we used the technology probe to experiment with a range of datasets.

the set of features, functions, and interactions that best supported the tasks at hand. These features and interactions formed the basis for our initial design. The technology probe also provided a platform for gathering initial informal validation, e.g., in the form of "I've been asking for this for 20 years!" During this phase, we also connected with the lead researchers of the two primary astronomy visualization and analysis packages (yt and Glue) to explore the possibility of developing our system as package extensions.

6.4.3.1 Artifact

The primary artifact of this phase was a technology probe, which we iteratively designed and developed, and which heavily informed our initial design. Screenshots of four iterations of the technology probe are presented in Fig. 6.2.

6.4.3.2 Reflection

As we found in our previous design studies [1],[3], the technology probe proved instrumental to engaging our collaborators and to creating a conducive and productive environment for investigating and testing ideas, learning about the domain and about our collaborators, and eliciting feedback and a broad range of ideas. The technology probe also allowed us to explore potentially more interesting visualization research opportunities, one of which was the use of interactive 3D scatterplots.

Interactive 3D scatterplots play an important role in our collaborators' research. They have been critical to discovering fundamental features and planes in galaxy observation data, and there is current interest in using them to look for more complex coherent structures and surfaces. As one collaborator put it: "I've built my career analyzing 3D scatterplots." These plots, however, are inherently problematic. They introduce issues of occlusion and spatial ambiguity, among other things. Approaches within the visualization community often turn instead to 2D projections of three+dimensional space, sometimes using animated transitions showing the original higher dimensional space in order to provide additional context [198], [199]. Building on existing work in this space [200], we see this collaboration as an excellent opportunity to further investigate the value of 3D scatterplots for analyzing and detecting features in three+ dimensional data. Is there anything that we can learn from this domain?

An underlying question throughout this design study was why the system had not already been developed. Our design was an obvious approach using thoroughly-established visualization techniques. When we brought up the question in a discussion with three faculty collaborators, they remarked that designing and implementing such a system was not an appropriate project for an astrophysics graduate student. When asked why the feature had not been implemented in existing analysis packages, they said they did not know. Even though this was the extent of inquiry, we see this collaboration as an excellent opportunity to explore potential barriers of adoption faced by domain researchers and practitioners surrounding the use of established visualization techniques. It would also be interesting to investigate how primary analysis packages are developed, what features are incorporated, and why and how researchers approach using these packages. How much of their analysis, and thus research, is guided by the capabilities of existing packages? Are there ways to design these packages that better support user-guided analysis? Can visualization research learn from, or contribute to, this area?

6.4.4 Iterative Design and Development

The final phase of the study was devoted to implementing our design. Our goals for this phase were: 1) to translate our findings and the results of our technology probe into a usable, deployable visualization prototype; and 2) to evaluate the prototype with domain experts via several different case studies. This phase spanned the last 4 months of the first stage of the study.

During the first 3 months of this phase, we met remotely on an either weekly or biweekly basis with one primary collaborator who was a postundergraduate researcher. Over the course of the meetings, we iteratively designed, developed, and refined the prototype using a case study that was central to the collaborators focus of research. During the final month, we met remotely with two collaborators, the postundergraduate researcher and our primary faculty collaborator, to prepare for and later to conduct the two case studies presented in Section 6.7.

During this phase, we also further explored the possibility of developing the prototype as extensions to the two primary astronomy visualization and analysis packages, yt and Glue. We met remotely with each of the lead researchers of these packages to discuss a development plan, and we attended and participated in a developers' workshop for yt. The workshop allowed us to investigate the ways in which our system could be meaningfully incorporated into the existing yt framework. It also provided an opportunity to interact with other astronomers and theoretical astrophysicists working in similar areas. These interactions allowed us to triangulate our study findings, gather informal feedback and validation, gather new case studies, and connect with potential future collaborators/users of the system.

6.4.4.1 Artifacts

The primary artifact for this phase is the implementation of our design in GalStamps, presented in Section 6.6.

6.4.4.2 Reflection

A significant amount of time during this phase was spent wrangling and preprocessing data for one of the case studies. Although this is a reported pitfall of design study [13], we have found that some degree of wrangling and processing is often necessary when visualizing new kinds of data. Rather than viewing this as necessary overhead, we wonder if there are ways that we can leverage these efforts for our research. For example, are there patterns in wrangling and preprocessing that we can characterize and provide guidance around?

Once we had finished wrangling and preprocessing the case study data, it became evident that our collaborators needed slightly different 2D scalar field data in order to investigate their initial research question. This was not a surprising result. Like many domain researchers, our collaborators often take an iterative approach to answering their research questions — trying something, seeing if it is informative, and then either iterating on the approach or trying something else entirely. This approach presents a challenge for design study, in which the goal is often to use visualization to help domain experts ask new questions of their data, or see their data in a new way. How can we support this iterative refinement of the question while also evaluating and validating our results?

6.5 Data and Task Abstraction

Within our collaborators' analysis, data items correspond to individual **observations** and datasets correspond to **collections of observations**. Whether real or simulated, a single observation corresponds to a region in the sky where a number of measurements are taken. These measurements are recorded as images and represented as a set of 2D scalar fields, where each field stores a different measurement, or *visual attribute*, across a 2D grid. Measurements derived from the set of 2D scalar fields are recorded as *statistical attributes* of the observation and stored as a row of tabular data along with the statistical attribute values of all other observations within the given collection. Thus, each observation in a collection is represented both as a set of 2D scalar fields and as a row of tabular data. The two representations are linked via an observation ID.

The goal of our collaborators' analysis is twofold: to verify that their theoretical simulations, and thus the underlying physics, are accurately capturing what is being observed in nature; and to use these simulations to help characterize what is being captured by observations. To meet this goal, our collaborators compare real and simulated galaxy observations. The associated visualization tasks involve *identifying* important features within each collection of observations and *comparing* these features across collections [201].

Identifying features within a single collection of observations involves analyzing statistical attributes within the tabular data, visual attributes within the 2D scalar data, and the relationships between these attributes. Visualization tasks include *locating*, *exploring*, and *browsing* [201] patterns, trends, correlations, and outliers within and across statistical and/or visual attributes. Comparing features across collections of observations involves analyzing the relationships [202] (similarities, differences, correlations, etc.) between features across datasets. Designing for this comparative level of analysis is a focus of future work in our ongoing design study.

6.6 Visualization Design

Our second contribution is a visualization design for statistically and visually analyzing real and simulated galaxy observations. We implemented our design in a prototype called GalStamps, named after the *galaxy postage stamps* that our collaborators analyze. The GalStamps interface, shown in Fig. 6.3, comprises a plot menu and two linked views: a plot view and an image view. Our design was grounded in the visualization literature and the results of our design study, and was heavily informed by prior work surrounding combined statistical and visual analysis using linked views [203], [204]. GalStamps was implemented in D3.js.

6.6.1 Plot Menu

The plot menu, shown as the leftmost pane in Fig. 6.3, allows users to generate custom plots from the statistical attributes available for a given dataset. We found manual selection of plot attributes to be an efficient approach, as each domain expert was interested in exploring unique and specific sets of attributes. Users can generate three types of plots: scatterplots, parallel coordinate plots, and binned-SPLOMs. We detail these plots in the proceeding sections.

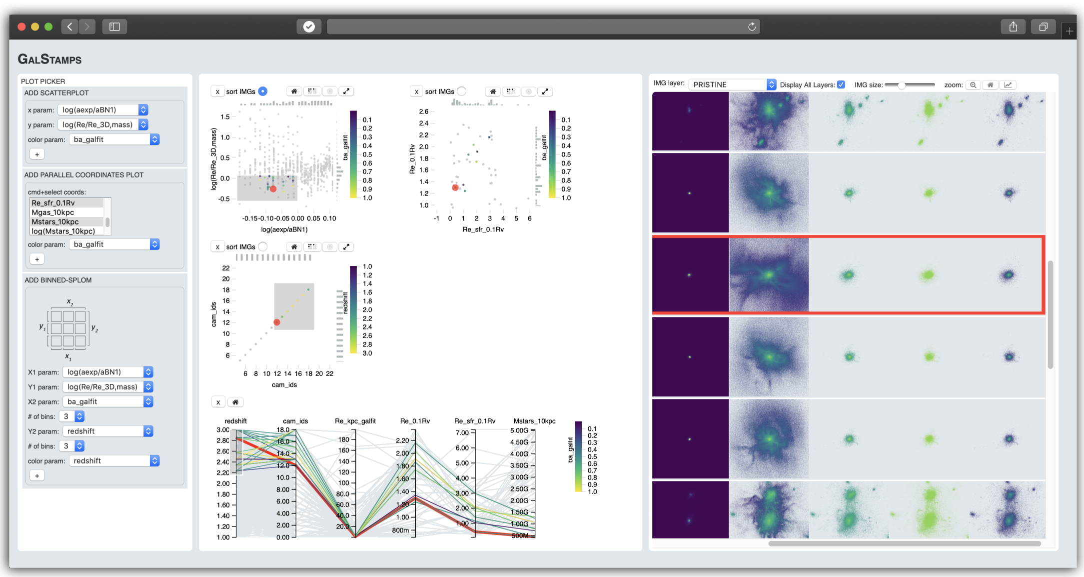

Fig. 6.3. The GalStamps interface. The interface supports the linked statistical and visual analysis of galaxy observation data. Statistical plots are generated via a plot menu *(left pane)* and displayed in the plot view *(middle pane)*. Selections within plots reveal the associated 2D scalar field data (rendered as images) in the image view *(right pane)*. Hovering over images in the image view highlights the associated statistical data in the plot view (shown in red).

6.6.2 Plot View

The plot view, comprising the middle pane of the prototype shown in Fig. 6.3, serves two primary functions. First, it supports the analysis of statistical observation attributes. As described in Section 6.5, this analysis involves identifying patterns, trends, correlations, and outliers within and across statistical attributes. We support these tasks through the analysis and comparison of plots. Second, the plot view provides a selection mechanism for querying 2D scalar field data based on selections of statistical attribute data. Selections are made via brushing. The plots are linked, such that selections in one plot are propagated to all other plots. For selections in multiple plots, the intersection of the selections is utilized.

Examples of scatterplots and parallel coordinate plots are shown in Fig. 6.3. In the **scatterplots**, observations are encoded as points of uniform radius, and histograms encode the distribution of points along each axis. Users can toggle to display isodensity contours. Point selections are made via a single 2D rectangular brush; isodensity contours are selected via mouse-click. Scatterplots support zooming and panning, and may be expanded to the width of the plot view. In the **parallel coordinate plots**, observations are encoded as joined line-segments, and selections are made via 1D rectangular brushes along one or multiple axes. Users can zoom and pan along each axis and reorder axes via dragging. Our initial inclusion of parallel coordinate plots was exploratory, as our collaborators had never seen them before. Interestingly, the domain experts predominately used the parallel coordinates plots as interactive legends for filtering high-dimensional data, which echoes the findings of others [205]. As described in Section 6.7, they also proved useful for summarizing the statistical attributes for a given observation.

Examples of the third type of plot, which we have termed **binned-SPLOMs**, are shown in the plot view panes in Fig. 6.4. In a binned-SPLOM, the bottom (x) and left (y) axes are preserved across all cells. Data in each cell are filtered into 2D bins, defined by secondary, top (x'), and right (y') axes. The static form of these plots is gaining popularity in the astronomy and astrophysics research community as an effective way to identify trends along pairs of secondary attributes [206], [207]. In this work, we experimented with adding interactivity to these plots. Zooming and panning allows users to adjust the ranges of the primary x and y axes. Histogram-scented filters allow users to adjust bin sizes and ranges along the secondary x' and y' axes. Observations are encoded as points, users can toggle to display isodensity contours, and selections are made via a 2D rectangular brush, or via mouse-clicks for isodensity contours.

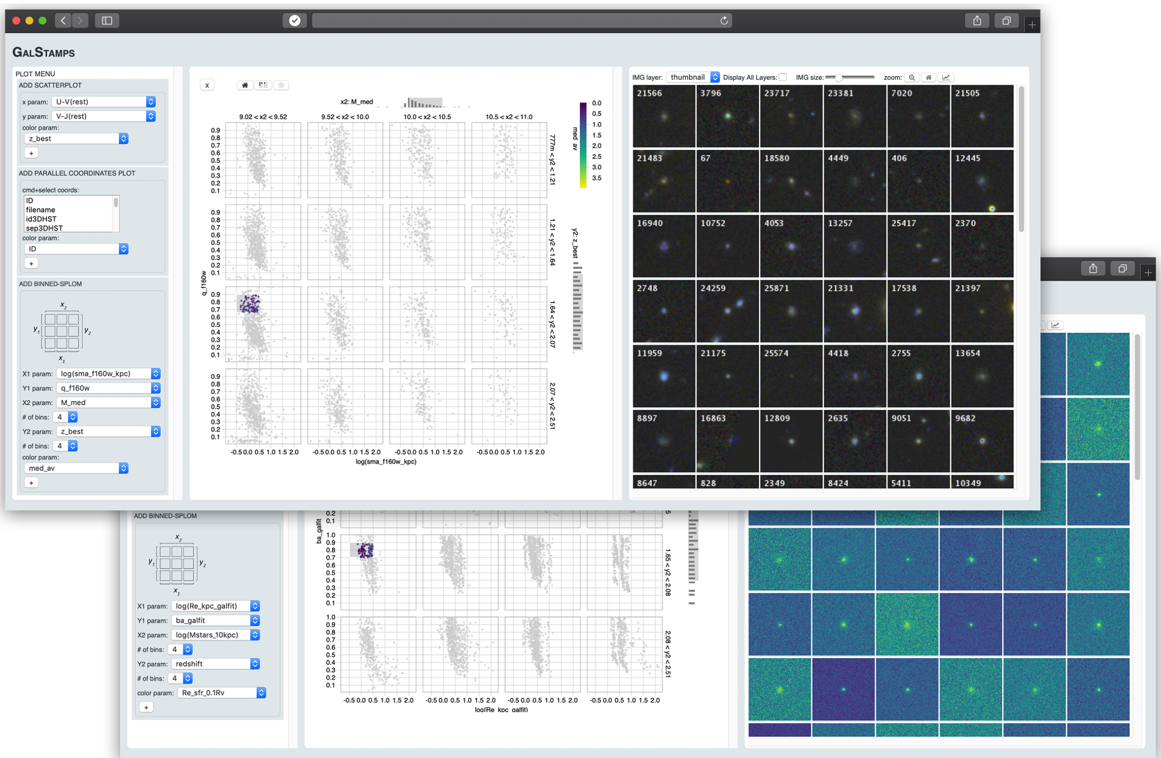

Fig. 6.4. GalStamps in practice. Our first collaborator used the prototype to compare real *(foreground)* and simulated *(background)* galaxy observations and to confirm prior claims based on deep learning analysis.

6.6.3 Image View

The image view, shown in the rightmost pane in Fig. 6.3, supports the analysis of visual attributes for a collection of observations. This analysis includes tasks for analyzing individual observations, represented as sets of 2D scalar fields and comparing multiple observations.

As a preprocessing step, 2D scalar fields are assigned a colormap and rendered as images. For a given observation, users can switch between different 2D scalar fields, via a dropdown menu, or can view the set of 2D scalar fields side by side. Image size can be adjusted via a slider, and users can toggle in and out of *zoom mode*, which supports zooming on images via mouse-scroll. Zooming on one image translates to all other images. For a given selection in the plot view, the associated images are displayed in rows and are ordered to reflect the *xy* distribution of the selection (the *selection space*) in the plot view. We liken this to an abstracted version of image scatterplots [208]. For intersected selections, the selection space defaults to the latest selection. Given multiple plots in the plot view, users can toggle between selection spaces via radio buttons. Images can also be reordered manually via dragging, which we found to be useful both for comparing images side by side and ordering groups of images based on visual patterns.

6.7 Case Studies

In this section, we present two case studies used to evaluate our initial design and its implementation in GalStamps. These case studies validated our integrated approach to enhancing and accelerating our collaborators' analysis and verified that our initial design was effective in supporting this approach. Additionally, the first case study revealed the need to support comparison of multiple collections of observations; the second case study illustrates the use of parallel coordinates for summarizing distributions along each attribute.

6.7.1 Comparing Real and Simulated Galaxy Observations

The first collaborator used the tool to examine findings from a recent paper [207]. In the paper, the authors employed data derived from both real and simulated galaxy observations to model the evolution of galaxy shapes within the CANDELS galaxy survey [209]. This collaborator began by examining a *real* collection of galaxy observations. Using the plot menu, he recreated a central figure, a binned-SPLOM, shown in Fig. 6.4 (*forefront*). In the plot view, he adjusted the histogram-scented brushes to reflect the original binning and began brushing to select various regions of interest within the different cells. With each selection, he studied the resulting images, hovering over individual images to reveal and assess the corresponding plot point. He was reassured that the images confirmed his expectations, and that he could reason about any disagreements.

In a separate browser, he then recreated and examined the same binned-SPLOM, this time generated using a *simulated* collection of galaxy observations, shown in Fig. 6.4 (*background*). Selecting the same regions of interest in the plot view, he confirmed that the associated simulated galaxy observations appeared as he expected. He commented that being able to easily pull up and inspect the associated images was useful. As a final step, he examined the two browsers side by side, selecting the same regions in each of the binned-SPLOMs and visually comparing the resulting sets of images. In each case, the simulated images looked very similar to the observed images, thereby validating the findings [210]: "We made this claim [...] based on deep learning analysis, but this is much better confirmation! [...] looking at the images and seeing how similar they are is really quite nice!" In discussing plans to support the comparative analysis of multiple datasets (e.g., real and simulated), he responded, “Why make things more complicated? This is fine!” Finally, this collaborator remarked on the broader utility of the system. Speaking about a related project, he commented, “tools like this one would be enormously helpful in analyzing these data.”

6.7.2 Deconstructing 2D Scalar Fields

The second collaborator intended to use the system to investigate a puzzling statistical feature in a simulated observation dataset. Questions about the image data, however, caused her to redirect her analysis, using the system instead to identify the visual attributes stored within each of the 2D scalar fields. By plotting related statistical attributes, selecting specific regions, and cross-verifying the associated images, she was able to test various hypotheses and eventually draw a tentative conclusion. A snapshot of this process is depicted in Fig. 6.3. Building on her conclusion, she was then able to pursue her initial investigation, generating preliminary hypotheses.

In reflecting on this process, she commented that the system enabled her to ask and answer her questions quickly and efficiently, in what would otherwise be a long, iterative process involving Python scripts. She remarked that seeing the tabular and image data together allowed her to explore the data in new ways. She also commented on the usefulness of the parallel coordinate plot for summarizing the statistical attribute values for a given observation, and for cross-verifying them with the images: “Instead of jumping around from plot to plot, I can just look at how each of [the attributes] are related.”

6.8 Discussion

The case studies presented in Section 6.7 provide validation for our data and task abstraction and our initial design and illustrate two ways in which the design can be used to enhance our collaborators’ analysis. Our collaborators found the GalStamps interface to be intuitive, and that the workflow — iteratively generating plots, selecting regions of interest, and cross-examining the associated observations — allowed them to effectively and efficiently pursue existing questions of their data, while also prompting new kinds of questions. These results give us confidence to continue with our initial design and reveal interesting design considerations moving forward. First, although our ultimate goal is to support the comparative analysis of real and simulated galaxy observations, and although our findings throughout the study indicate the importance of designing for this level of comparison, we wonder whether utilizing two side-by-side versions of the system is in fact sufficient, as evidenced by our first case study, or whether there is significant added benefit integrating this analysis. Second, feedback gathered with the technology probe and in the second case study suggests a different use of parallel coordinate plots beyond exploring correlations between attributes: as an interactive legend for filtering and summarizing high-dimensional data.

Our findings throughout the first stage of this design study point to two interesting mismatches between visualization guidelines and domain practice. First, as we describe in the rich process description included in supplemental materials, interactive 3D scatterplots play a valuable role in our collaborators' current analysis workflow. As one collaborator put it: "I've built my career analyzing 3D scatterplots." Three-dimensional scatterplots, however, are argued against within the visualization community [198], [199], [211]. Building on existing work in this space [200], we see this collaboration as opportunity to further investigate the potentially underappreciated value of interactive 3D scatterplots for analyzing and detecting features in three+dimensional data. Second, an underlying question for us throughout this study was why the system had not already been developed. The GalStamps prototype uses a standard multiple linked view approach with rich interactions, but these design standards have not yet made their way into the astrophysical community. We see this collaboration as a valuable opportunity to investigate the barriers that domains face in adopting established visualization techniques.

6.9 Reflection

In what follows, we reflect on how our ADR-informed actions strengthened our early-stage design process and results. We additionally discuss the important questions that emerged and the results of experimenting with new ways of supporting ADR throughout the design process.

6.9.1 Further Validating ADR

The results of this design study further validate ADR as a useful model for strengthening elements of the applied visualization research process. As with our previous design study, ADR's emphasis on research context played an important role in guiding the design process and development of results. In this study, *really listening to our domain collaborators* — to their desires, hesitations, and perspectives — led us to design a visualization platform that filled an important gap in our collaborators' analysis, allowing them to interrogate, validate, and correlate their data in new and more efficient ways. Our initial results give us confidence that our platform will be adopted and used long-term, especially provided the necessary technical support.

In addition, attending to the dynamics of the research team helped us collaborate more effectively throughout the study. Our team comprised senior researchers (faculty) and junior researchers (graduate students, undergraduate students, and research assistants). Our meetings involved various combinations of team members, from one-on-one meetings with senior or junior researchers, to group meetings with senior and/or junior researchers, to team-wide meetings. Studying the dynamics of the different collaborative environments, including our influence as visualization researchers, allowed us to leverage our meetings for different kinds of learning. This was a strategy that we experimented with in the previous design study but were able to implement more fully in this study. We learned to use our meetings with senior researchers to discuss science questions and relevant background, triangulate ideas, and map out research paths. We used our meetings with junior researchers to translate science questions and formulae into raw and derived data, discuss low-level design requirements, and formulate use-case scenarios. Finally, we used our meetings with senior and junior researchers, as well as team-wide meetings, to understand the various stages of our collaborators' research pipeline and how these stages could be supported within a visualization platform. Our experience contributes to discussions within the design study community about collaborator roles in an academic setting [8].

ADR once again provided a valuable structure for our learning and for the formalization of our learning. We framed our learning around cycles of intervention, with each cycle building on the results of the last. We employed a technology probe [31] as a central mechanism for intervention. Cycles were interspersed with critical reflection, which was facilitated through our continued practice of taking field notes. Our reflection spanned the challenges we encountered, strategies we adopted, forms of validation we received, shifts in our research thinking, shifts in our collaborators' research thinking, broader implications of our work, and important future research directions. As supplemental material to the publication of the study, we include a rich process description framed around the different cycles of intervention and based heavily on our field notes. For each cycle, we provide the details of our process and the resulting artifacts and a synthesis of our critical reflection. In the design study with global health experts, this rich process description was incorporated into the main manuscript. The content was heavily synthesized and polished, presenting only the details that helped validate aspects of the study. In contrast, including the rich description instead as supplemental material afforded more space to include questions, speculations, and details that did not directly validate the study, but that could prove useful to other visualization researchers. The resulting description is less synthesized, less polished, and arguably more open and authentic in its account of the research process. We see different value in each level of description. We speculate that including both descriptions, or some combination of the two, would maximize the impact of this contribution.

Our use of a technology probe once again proved helpful for engaging our collaborators and establishing a conducive environment for brainstorming, experimenting, and eliciting feedback. In this study, we found it more useful to employ a single, evolving technology probe rather than multiple isolated probes. Over the course of the study, the probe evolved to support our growing understanding of the domain problem and the data and tasks. Early iterations were used to explore the problem space, to probe our collaborators' relationship with data and visualization, and to demonstrate the extent of what was possible. As the study progressed, we used the probe to ask more specific questions and test specific design decisions. Having the ability to experiment with interactions and encodings in real time with our collaborators helped us establish credibility as visualization designers and give our collaborators agency to promote their own ideas, and also allowed for tight cycles of collaborative iteration and development. In addition, what we began to observe in our two previous studies, and what became particularly clear in this study, was that involving our collaborators at this level of development increased their sense of ownership of the ultimate design, while also promoting a view of the design as dynamic and evolving rather than static and final.

Finally, ADR, and our experience and success applying ADR in the previous design study, conditioned us to pursue and formalize themes that emerged out of moments of disruption to our visualization research thinking. The first theme, examining barriers to adoption of established visualization techniques, stemmed from our amazement that the system we were designing had not already been developed, despite being such a clear application of standard linked views. The second theme, investigating the utility and value of techniques that the visualization community in large part recommends against, stemmed from our domain collaborators' validated use of interactive 3D scatterplots. These themes represent mismatches between visualization guidelines and domain practices. Although not central to our early-stage results, we offer these themes as potentially important directions for future visualization research.

6.9.2 Emergent Questions

Several important questions emerged from challenges we encountered when implementing our ADR-informed actions. Our context-driven approach led us to pursue a design that offered significant value from a domain research perspective, but little value from a visualization research perspective. In reflecting, we ask: Did we perhaps listen *too* carefully to our collaborators? Should we have pushed harder on our visualization research agenda at the risk of designing something less relevant to our collaborators' research? Or, should we have abandoned the project, deeming it an unsuitable design study [8]? An underlying question throughout this study was how to conduct strong applied visualization research when there is a clear, conventional visualization solution, and when deriving new visualization knowledge is at odds with designing effective solutions to real-world problems. We suspect that this tension extends beyond visualization to HCI and to applied design research more broadly. Our approach to addressing this conflict was to use the conventional visualization solution to probe for more theoretical, visualization contributions.

This approach did uncover interesting visualization research questions surrounding mismatches between visualization guidelines and domain practice — questions that we may not have encountered otherwise and that we began to investigate over the course of the design study. These questions, however, were not the focus of our publication, which we instead framed as an application of standard link-views. When it came to presenting our results, we struggled to formalize our learning around these mismatches in a way that would be valuable for the visualization community. This experience led us to ask how we might formalize preliminary learning such that we or other visualization researchers can build on this learning in future research, and how we might curate our raw field note data to better support this. Along similar lines, in publishing the results of this study and our previous study with global health experts, we offer a rich process description and raw field note data in an effort to increase the reliability and transferability of our research. We have yet to understand, however, whether these materials are in any way effective. Do rich process descriptions and raw field note data help increase the reliability and transferability of research? Can we present these materials in a way that better supports this? One approach to answering these questions involves evaluating these materials with other visualizations researchers, analyzing the results to derive preliminary guidelines, and testing and refining the guidelines in follow-up studies. Our descriptions and raw field notes offer an initial dataset for this evaluation.

Lastly, as part of our reflection and learning, we set out with the intention to reflectively transcribe all recorded interactions with our collaborators. Reflective transcriptions had proven valuable in previous studies for capturing recorded moments of insight and generating and capturing new insights, and we had hoped to further test and validate this mechanism in our study. The nature of our collaboration, however, made this difficult to implement in practice. Throughout the study, we relied heavily on semistructured and unstructured discussions, giving our collaborators ample space to present their perspectives. As a result, many meetings turned to lengthy discussions of the domain. These discussions provided valuable opportunities to observe the domain research in action and the dynamics among team members, but presented challenges when generating reflective transcriptions from hours of recordings that were in large part only loosely relevant. This challenge, combined with the fact that the collaboration was structured into visits filled with back-to-back meetings, made generating reflective transcriptions a particularly daunting and questionably time effective activity. Ultimately, only a handful of these transcriptions were completed. We ask whether there is a more reasonable way to approach this. One simple strategy might be to take note of times in which relevant discussion is taking place and to limit the reflective transcriptions to only those times. This strategy, however, is not without issues. For example, it runs the risk of skipping over important insights or disengaging the visualization researcher from relevant discussion. We suspect that more rigorous strategies exist or may be developed.

6.9.3 Experimental Results

In our design study with global health experts, field notes emerged as a valuable mechanism for facilitating critical reflection and for generating a trail of evidence of the design process and development of results, which we published in an interactive timeline alongside the design study. Building on these results, we used the current design study to experiment with new ways of enhancing and supporting field note taking throughout the visualization research process. Our results point to interesting areas for future visualization research.

To experiment with enhancing field notes, and drawing field research methods, we incorporated an additional practice of distinguishing between descriptive field notes, comprising factual data about context and interactions, and reflective field notes, comprising questions, interpretations, and ideas [164]. We anticipated that this added practice would encourage us to be more purposeful in our note taking, capturing moments of learning as well as the context in which the learning occurred. We also anticipated that, when it came to formalizing and presenting our results, explicitly including and differentiating between factual and interpretive field note data would prove useful for tracing the origin and evolution of ideas and for substantiating claims.

We planned to be diligent in our practice, distinguishing between the descriptive and reflective content when taking field notes in meetings, and fleshing out and analyzing our notes directly following meetings [164]. In reality, however, sticking to these plans proved difficult given the densely packed, highly participatory nature of our meetings. Although we found that we could, on occasion, code our notes as descriptive or reflective in real time, in most cases we found ourselves returning to the notes to differentiate and analyze them retrospectively — no doubt diminishing the intended impact of the practice. We suspect that this difficulty points to important differences between field research, in which the researcher is an observer, and applied visualization research, in which the researcher is a driving participant. In going back through our field notes to formalize our process and results, we found that our notes ultimately did a better job than in the previous design study of capturing the context surrounding moments of learning and the rationale behind decision-making. The notes were less effective, however, in capturing the origin and evolution of ideas and insights. Adapting this practice to better suit our research and better supporting this practice throughout the research process are two valuable directions for future work.

To experiment with supporting field note taking throughout the research process, we developed a *visualization research* technology probe, which we employed over the course of the design study. The probe was based on the interactive timeline developed in our previous study, which we found to be a powerful and intuitive way to view and explore field note data, but which was not supported by existing note-taking applications. The technology probe, shown in Fig. 6.5, was developed using D3.js and node.js and functions as follows: field notes comprising text and media (images and audio) are submitted via a mysql form (left panel). Once submitted, notes are added to an interactive timeline (right pane). The timeline supports filtering via scrolling or brushing. Individual field notes can be viewed and edited via hovering or clicking on thumbnails in the timeline. Field note data can be backed up manually and the timeline can be exported to a static version that can be published online for example, as supplemental material for a publication.

The probe allowed us to explore the problem and solution spaces; to develop an initial data and task abstraction for supporting field note taking, analysis, and dissemination; and most importantly, to gain a general sense of the potential value of such a platform. In using the probe, we found ourselves alternating between entering field notes directly into the interface, cutting and pasting

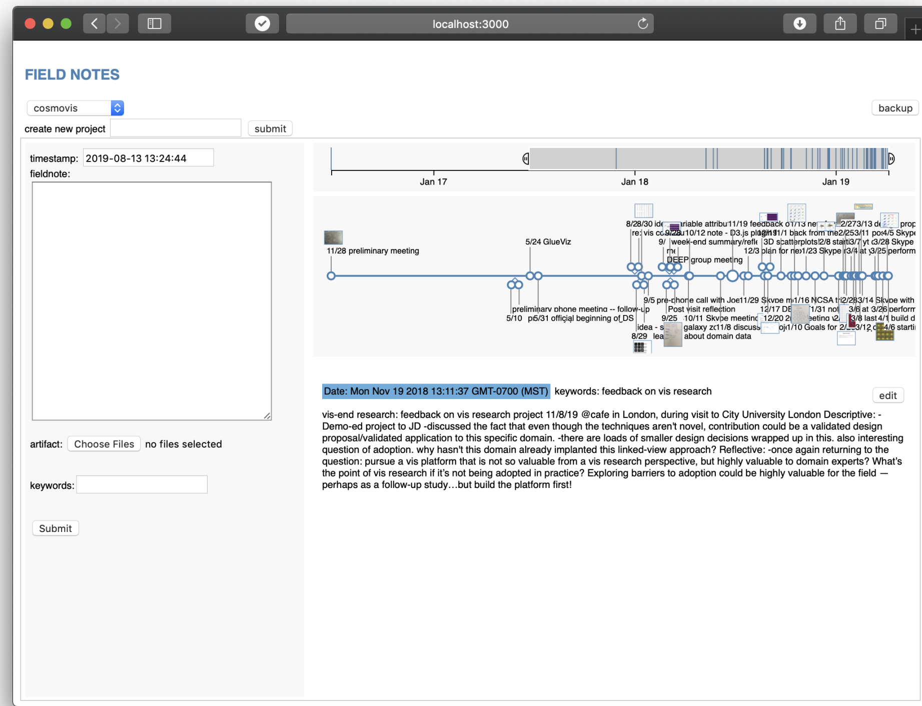

Fig. 6.5. *Visualization research* technology probe. To experiment with supporting field note taking throughout the research process, we developed a *visualization research* technology probe, which we employed over the course of the design study.

our notes from other text editors, and transcribing them from written notes. We documented our feedback on the features and functionalities and developed the tool, when possible, to better support our process. Throughout the design process, we found that the probe encouraged us to take field notes more frequently; to recount seemingly inconsequential events and interactions; and to jot down questions, ideas, and insights that we previously would have stored mentally and potentially lost. When preparing our results for publication, the interactive timeline provided an effective and engaging platform for exploring and analyzing field notes across the entire study. This platform proved very useful for formalizing and synthesizing our process and for extracting and contextualizing important themes and questions that emerged.

Ultimately, our results suggest that a field note taking platform has the potential to greatly enhance reflection and documentation throughout the design process and to facilitate the formalization and dissemination of learning. This result is contingent, however, on the following design requirements: first, the interface for submitting field notes must be easy and intuitive to use; ideally as intuitive as taking notes in a notebook. We view this requirement as most critical to the successful adoption of the platform. Second, and along the same lines, a user must have the ability to alternatively import field notes from his or her preferred editor. Third, the interface should center around an interactive timeline. We see the timeline as a powerful component that distinguishes the platform from other note-taking tools. Fourth, the platform must export a public-facing version to be published alongside the results of a study. What this public-facing version looks like, and how to support the exploration of field note data by external audiences more broadly, presents a rich space for visualization research and design. Finally, the tool must support manual and/or automatic backup of field note data.

Our results also suggest that such a platform could be further designed to address many of the challenges we confronted when applying ADR to visualization design — for example, by integrating mechanisms to better support taking reflective and descriptive field notes, generating reflective transcriptions, and formalizing preliminary learning. Overall, we view the design and development of a field note taking platform as an excellent visualization-for-visualization-research design study with powerful implications for applied visualization research. We hope that our work can help promote and inform future research in this area.

CHAPTER 7

REFLECTIVE SYNTHESIS OF EXPLORATORY RESULTS

This dissertation explores the utility of ADR for navigating and leveraging the contextual, situated nature of applied visualization research. Our exploration was motivated by a design study with poetry scholars, presented in Chapter 3 and informed by preliminary theoretical research on the ADR framework, presented in Chapter 2 and Chapter 4. We then applied and extended the ADR framework in two consecutive design studies — the first in collaboration with global health experts analyzing Zika outbreak data, presented in Chapter 5, and the second in collaboration with astronomers and astrophysicists studying galaxy formation and evolution data, presented in Chapter 6. We reflect on our experience conducting these studies at the end of each chapter.

In this chapter, we present a reflective synthesis of our exploratory results. Our results validate ADR as a useful model for strengthening elements of the applied visualization research process, while also revealing significant gaps in the ADR framework. These gaps pose important areas for future methodological visualization research. As part of our synthesis, we speculate on one way to address these gaps through better recording, reflecting, and reporting of artifacts in design study.

7.1 Validation of ADR

Our exploratory results validate ADR as a useful model for strengthening various elements of the applied visualization research process. At its core, ADR strengthens applied visualization research by conditioning researchers to recognize, attend to, and leverage the role of people and context in shaping the design process and the development of results. In the Zika design study, carefully attending to people and context proved critical to the success of the project, from navigating the field research environment, to defining a suitable collaboration, to negotiating the expectations of our collaborators and the domain expert community in which the design study was embedded. In the galaxy design study, this same careful attention allowed us to leverage the dynamics of the research team in order to collaborate more effectively. Although the influence of people and context underlies all collaborative research environments, ADR brings this element to the forefront, encouraging researchers to account for and leverage this influence in all aspects of the design process.

On a practical level, ADR provides a valuable framework for conducting applied visualization research in highly contextual, dynamic, and human-centered environments:

- **ADR provides structure for learning via cycles of intervention and disruption.** In ADR, these cycles are designated to support the iterative development of the technology artifact. In applying ADR, we found that framing all stages of research, from problem formulation to formalization of learning, as cycles of intervention and disruption helped to promote learning throughout the research process. In the Zika design study, major cycles were devoted to preconditioning and problem formulation, learning about Zika outbreak analysis, learning about data discrepancies, and formalizing our learning. In the galaxy design study, these cycles were devoted to preconditioning and problem formulation, learning about galaxy observation analysis, exploring the design space, and iteratively designing and developing a prototype. In both cases, as was the case in our poetry design study, cycles were mutliscaled, comprising supplementary investigations as well as rapid cycles of experimentation and feedback. Although these cycles reflect the general stages and phases of research defined in existing design study methodologies [7], [13], we found the more focused framing around specific points of learning to be particularly helpful in guiding our research trajectory.

  ADR's cycles of intervention and disruption are guided by the principle of *reciprocal shaping*, which emphasizes the constant shaping of design process by the different perspectives within the team, and the principle of *mutually influential roles*, which emphasizes the learning that should occur on both sides of the collaboration. In both the Zika and galaxy design studies, adherence to these principles pushed us in interesting and valuable directions with our research. ADR's principle of reciprocal shaping encouraged us to *really listen to our domain collaborators*, while also listening to our own needs as visualization researchers. In the Zika design study, the interplay between our collaborators' ambivalence toward data visualization for analysis and our desire to be data visualization evangelists became the driving factor for our research. In the galaxy design study, reciprocal shaping manifested as a constant negotiation around pursuing research that was valuable to both the domain experts and the visualization researchers. ADR's principle of mutually influential roles, on the other hand, gave us confidence to pursue insights stemming from moments of disruption to our and our collaborators' research thinking. In the Zika design study, the resulting framework stemmed from re-evaluating the role of visualization for analysis. In the galaxy design study, adherence to this principle led us to an investigation of mismatches between visualization guidelines and domain practice. Both principles conditioned us to be attuned to and guided by things that surprised us and by moments of friction with our collaborators. In applied visualization research, these instances present rich opportunities for learning and insight.

- **ADR provides structure for enhancing and leveraging learning through continuous and concurrent critical reflection.** In our initial poetry design study, reflection at the end of the study led to important discussions surrounding the role of disruption, breaking visualization convention, and appropriate measures of success. Had we reflected on these elements earlier in the study, we could have leveraged the study context to pursue these lines of inquiry more fully. This missed research opportunity motivated us to look to fields outside of visualization, and eventually to the ADR framework, for guidance on how to reflect more purposefully throughout the design process. Throughout the Zika and galaxy design studies, ADR encouraged us to critically reflect on the evolving design process, our interactions with our collaborators, and our moments of insight and learning, and to use our reflection to help guide the research process. In the Zika design study, critically reflecting on our early stage research, and specifically on our collaborators' hesitation to use data and visualization for analysis, led us to shift our research in a direction that proved more interesting and valuable to both ourselves and our collaborators. In the galaxy design study, continuous critical reflection led us to branch out, pursuing multiple lines of visualization research inquiry in parallel with our primary design process. When compared to the poetry design study, we argue that this continuous, critical reflection enhanced our learning throughout each of these subsequent design processes.

- **ADR provides structure for formalizing learning into generalized visualization knowledge.** In both the Zika and galaxy design studies, we revisited our reflection throughout the design process in order to synthesize and formalize our learning into broader visualization knowledge. In the Zika design study, we devoted an entire phase of research to the formalization of our learning, resulting in the primary contribution of the study: a conceptual framework. In the galaxy study, we formalized our learning surrounding mismatches between visualization guidelines and domain practice into important areas for future visualization research. Over the course of the two studies, we additionally came to view the formulation of research questions as a valuable outcome of the formalization process. Our experience points to the extent of generalized knowledge that can result from design study, from major theoretical contributions to research questions and directions for future work. The confirmation, refinement, rejection, or proposal of visualization guidelines are other examples of resulting generalized knowledge, as are problem characterizations and data and task abstractions [13]. ADR helps to increase the level of resulting generalized knowledge, first by encouraging researchers to reflect on the broader implications of their learning *as it progresses in real time*, and second by encouraging researchers to take the time to revisit, synthesize, and formalize their reflections.

- **ADR provides structure for cross-study analysis.** Although perhaps not deliberate, we found that ADR facilitates cross-study analysis. In reflecting on our galaxy design study, we were naturally compelled to compare and contrast aspects of the study (the context, the different cycles of intervention, and the topics of reflection) with those of the Zika study as well as with those of the poetry design study, which we retrospectively framed through the lens of ADR. Whereas existing visualization models guide researchers through the different stages of the design process — from characterizing the problem, to developing a data and task abstraction, to designing and validating a visualization solution, to deriving broader visualization guidelines — ADR promotes comparison by encouraging researchers to step back and ask questions such as: How did the people and context of this study shape the process and results? What did disruption look like in this context? What were the different kinds of learning that occurred? Given the highly situated nature of design study, cross-study analysis offers a valuable alternative approach to creating generalized outcomes.

7.2 Gaps in the ADR Framework

Our exploratory results reveal significant gaps in the ADR framework. We frame these gaps as four important areas for future methodological research. First, ADR fails to adequately address the issue of reliability in applied visualization research. In particular, ADR sidesteps the challenging issue of reliability, and rigor, more broadly, in applied visualization research by resorting to the notion of *trustworthiness* — an alternative view on rigor for qualitative, situated design studies, which is now quite dated [212]. Second, ADR for applied visualization research would benefit from a shift in emphasis from generalized outcomes as the desired result of the formalization of learning, to *speculated* generalized outcomes and *transferable outcomes*. Applied visualization researchers who are deeply embedded in a research context are in the best position to speculate about how their results generalize to broader contexts. It should be stressed, however, that these are, in nearly all cases, speculations. Transferable outcomes, on the other hand, support the adaptation of results by other researchers to their specific research context by providing the necessary contextual detail. Generating transferable outcomes in addition to any speculated generalized outcomes has the potential to greatly extend the impact of applied visualization research.

Over the course of the Zika and galaxy design studies, we experimented with one approach to increasing both the transferability and reliability of our research through the inclusion of rich process descriptions and raw field note data alongside the publication of our study results. Our approach was grounded in methods from the social sciences and field research. These efforts contribute to the ongoing dialogue within the visualization community, but also indicate that much more research and standardization are needed in this area. First and foremost, we need to formally define what reliability and transferability look like in the context of applied visualization research. We then need to better understand what these criteria afford and how they are established, evaluated, and utilized in practice. From there we need to evaluate and refine our approach with other visualization researchers and in light of these definitions and this deeper understanding.

Third, ADR for applied visualization research would benefit from a broader notion of artifacts. In the Zika and galaxy design studies, technology artifacts — technology probes and system prototypes — constituted only one kind of artifact. Other reported artifacts included problem formulations, data and task abstractions, and resulting conceptual frameworks. These artifacts represent externalizations and formalizations of the learning that occurred and evolved throughout the studies. Broadening ADR's notion of artifacts to include these externalizations and formalizations of learning would better reflect the range of contributions that stem from design study [26] and would better support capturing the development of these contributions throughout the research process.

Fourth and last, ADR for applied visualization research would benefit from methods and mechanisms that explicitly support the stages and adherence to principles throughout the design process. We have proposed a number of mechanisms that emerged as useful for supporting ADR throughout our design studies: field notes, reflective transcriptions, technology probes, and a field note taking platform. We have also discussed some of the limitations of these mechanisms, for example the challenge of taking field notes in back-to-back, highly participatory, collaborative settings. These mechanisms need to be further developed and formally evaluated, and we anticipate that many other valuable mechanisms remain to be discovered, invented, and adapted.

In summary, our exploratory results reveal that ADR fails to adequately support reliability and transferability of applied visualization research, and would benefit from a broader notion of artifacts as well as from mechanisms that explicitly support ADR throughout the design process. Based on our learning over the course of three design studies and our theoretical research into the ADR framework, we speculate that these gaps may be addressed through better recording, reflecting, and reporting on artifacts in design study. Artifacts provide a platform for richly documenting the design process. Recording, reflecting, and reporting on artifacts offers a mechanism for supporting reliability and transferability in the context of ADR's stages and principles by increasing the transparency of the research process and design rationale. In the next section, we formalize our learning surrounding this potential strategy.

7.3 Recording, Reflecting, and Reporting on Artifacts in Design Study

Within the qualitative social sciences, as described in Chapter 2, reliability of research is established in part by systematically documenting research practices and the emergence of findings, and by allowing such documentation to be audited by other researchers [21]. Supporting transferability, on the other hand, entails supplying readers with a database of sorts containing the contextual details that enable transferability judgments by potential appliers. Both require rich documentation of the research process and development of results. This requirement poses a challenge in applied visualization research, in which the design process is emergent, contextual, subjective, and often extemporaneous.

We propose **artifacts** as a platform for documenting the applied visualization research process, and thus as a means for enhancing the reliability and transferability of the design process and findings. We broaden the term artifact to describe the physical and digital documents generated throughout the design process, from sketches and field notes, to recorded interactions, to prototype iterations. Artifacts are instantiations of the insights that drive the design study. They are externalizations of the learning and evolving understanding that guide the design process, inform design decisions, and ground the development of broader visualization knowledge. Additionally, artifacts inscribe the contextual, subjective environment in which they were generated. Analyzing artifacts can help explicate the role of context and environment in shaping the design study.

When captured as a sequence, these artifacts provide a record of the emergent design process. If properly analyzed and communicated, these artifacts could also provide an account of the learning that shaped the process and results, and the ways in which the process and results were influenced by, and subject to, the research environment. Including such a record in or alongside the publication of a design study could enable fellow visualization researchers to make their own judgments surrounding the reliability of the research and results, as well as surrounding the transferability of various aspects of the design study to their own research contexts. Despite this potentially powerful approach to enhancing reliability and transferability of design study research, there is limited discussion surrounding the range of artifacts generated in design study and how to effectively capture, analyze, and communicate these artifacts to the visualization research community.

We use the terms *recording*, *reflecting*, and *reporting* to describe what we have identified as three key activities for capturing, analyzing, and communicating design study artifacts. **Recording** refers to the externalization of learning as artifacts throughout the design process. **Reflecting** describes the process of analyzing and learning from captured artifacts and incorporating this learning into the design process and the development of results [8]. **Reporting** describes the process of transforming artifacts into transferable visualization knowledge and curating the results into an audit-able record of the emergent design process. We anticipate that these activities can also foster new kinds of artifacts and support the externalization of additional learning and insights throughout the design process.

In what follows, we explore a broader notion of artifacts that is grounded in the concept of artifacts established within the design community and the social sciences as well as in existing notions within the visualization community. We then present work surrounding existing approaches to recording, reflecting, and reporting. We conclude with a discussion of important areas for future applied visualization research.

7.3.1 A Broader Notion of Artifacts

Our strategy hinges on the broader notion of artifacts as externalizations and articulations of learning throughout the design process. The notion of artifacts is deeply rooted in design and can be traced back to the development of design science [213]. In design science, artifacts are defined as the outcomes of design that function within the environment in which they are generated [213], [214]. This conceptualization of the design artifact has been embraced and developed in many areas of design. Within information systems (IS) research, conceptualizations of the *IT artifact* have evolved from a lens for viewing and discussing the various roles and functions of IS research technologies [29], to constructs, models, methods, and instantiations [215]. More recently, members of the IS research community have embraced a more holistic view of the IT artifact, as embodying the theory, context, and learning that shape the design process [5], [30].

The notion of the artifact has also been embraced and developed within HCI. Research through design [216], a popular model for interaction design within HCI, presents a conceptualization of the design artifact that reflects and extends those developed in IS research. Artifacts are defined as outcomes of design research; they range from models, to prototypes and products, to documentation of the design process; they inscribe design knowledge and reflect a specific framing of the problem; and they are context-dependent and are posed as questions to the community. Within the research through design literature, artifacts are demonstrated as serving a number of different functions, from reflecting on a design process and design decisions, to promoting and refining existing methods, to generating design theory, to evaluating with end-users [217]. These functions reflect the contribution scenarios presented in the design study literature [26]. We return to these scenarios later on in this section. Additionally, research through design artifacts is presented as *constituting* knowledge, *containing* knowledge, and *constructing* knowledge [217]. The view of artifacts as platforms for knowledge is central to models of distributed cognition, which posit that knowledge and understanding of a problem are distributed across a collection of individuals and artifacts [218], [219]. Distributed cognition is promoted as a theoretical foundation for HCI methods [220]. A slightly more complex notion of artifacts has also developed within the research through design and broader HCI communities: *concept-things* [217], *strong-concepts* [221], and *annotation portfolios* [222], [223] all describe artifacts that combine design articulations (previously referred to as artifacts) and verbal articulations (textual descriptions of the knowledge embodied within the artifact).

Consistent throughout the design, information systems, and HCI literature is the idea that artifacts embody elements of the research — the theory, knowledge, learning, and context that shape and inform the design process. In our proposed strategy, we pivot slightly, framing the artifact instead as a platform for externalizing, contextualizing, and analyzing these critical elements of design study, and communicating them back to the visualization community.

Turning now to the field of visualization, early notions of artifacts already exist within the research community. In the nine-stage framework [26], the notion of artifacts is associated with the three types of design study contributions: problem characterizations and abstractions, validated visualization systems, and reflections of lessons learned. More recently, the range of possible design study contributions has been extended to include the results of seven contribution scenarios: propose a novel technique, reflect on methods, illustrate design guidelines, transfer to other problems, improve understanding of a subarea, address a problem that your readers care about, and provide a strong and convincing evaluation [26]. These scenarios offer insight into the range of artifacts that exist in design study.

In the design activity framework [7], the notion of artifacts is analogous to *outcomes*, which include everything from tangible software tools and sketches to ideas, users' needs, and lists of software requirements. Outcomes are described as the specific, unique result of design activities, which stem from making design decisions. Outcomes are promoted as a way to enable richer design process descriptions, to document design decisions, and to enhance actionability of contributions.

The concept of data sketches [32] offers another example of notions of artifacts in the visualization literature. Data sketches describe exploratory and evolving visual representations of domain data, captured over the course of the design process as sketches, wireframes, and paper and digital prototypes. Data sketches are shown to be particularly effective in facilitating the early stages of human-centered design, from developing an understanding of the application domain, to establishing design requirements, to developing early- and late-stage system prototypes.

Design study contributions and contribution scenarios, design activity outcomes, and data sketches all offer important insight into the broad range of artifacts that are generated in design study and that help to promote reliability and transferability. We anticipate, however, that these examples represent only a small subset of the range artifacts that are important for increasing reliability and transferability. Looking beyond the field of visualization for additional insight and guidance, we highlight three different research concepts — thick description, field notes, audit trails

— that exist outside the current notion of design study artifacts, but that we think could be valuable for enhancing transferability and reliability of research process and findings. We cast these examples as a narrow subset of the range of artifacts generated throughout the design study process.

The first example artifact is *thick description*. Thick description is a fundamental concept in qualitative research, promoted not only as a mechanism for transferability [20], [21], [224]–[227] as noted in Chapter 2, but also as a mode of research [228]–[232] and a mechanism for enhancing credibility of research [226]. Despite its widespread acceptance and use, however, definitions of thick description are elusive [233]. The concept is characterized more by what it *does* than what it *is* or *contains*. Thick description creates a sense of verisimilitude, enabling readers to place themselves in the experience and the context and to evaluate it against their own experiences and context in order to adapt the elements that seem relevant and useful [230]. Thick description is often contrasted with *thin description*, which contains factual statements that lack detail and density. For example, the wink of an eye may be described as a muscle movement (thin) or as a message with implications (thick) [234]. Thick description offers insight into how a reader might go about identifying potentially transferable findings. It also, however, points to the commitment required by the researcher and the reader to make the transfer happen. Making transferability worthwhile is a critical area for future applied visualization research.

The second example artifact is the *field note*. Field notes are an integral part of qualitative fieldwork. The term broadly describes notes taken by researchers during fieldwork in order to record and remember observations and interpretations. Field notes are intended as evidence to support claims about meaning and understanding [164], [235]. They generally contain two kinds of information: descriptive information and reflective information. *Descriptive information* contains factual data about context, actions and interactions, perspectives, and influences. In contrast, *reflective information* records thoughts, ideas, and questions that emerge during an observation. Field researchers are encouraged to analyze their field notes as they are being written and during observation, as this fosters self-reflection and can reveal emergent themes that can inform the investigation [164].

The third example artifact is the *audit trail*. Based metaphorically on financial audits, audit trails support the careful inspection and verification of research process and findings and are promoted as a mechanism for enhancing reliability in qualitative and subjective research contexts [21]. Audit trails contain evidence of reported actions, decisions, and findings. Such evidence is collected throughout a study, compiled chronologically, and made available for scrutiny by an external audience. Audit trails can exist as single artifacts [12], for example as a textual narrative, or as a compilation of many artifacts [11]. The notion of audit trails is just beginning to gain traction within the visualization community. Litvis [12] explores approaches to documenting design rationale, in real time, throughout the design process. The Litvis system supports annotating visualization implementation code with textual narrative describing design rationale. This work seeks to improve accountability of visualization research via the rich, curated documentation of design process and findings. Additionally, recent methodological work focusing on the use of creativity workshops in design study includes, as supplemental material, a comprehensive audit trail of the research process [11]. The audit trail presents a compilation of 30 artifacts documenting the critically reflective practice that guided the research.

7.3.2 Recording, Reflecting, and Reporting

We identify recording, reflecting, and reporting as three key activities for capturing, analyzing, and communicating design study artifacts. Our understanding is grounded in the visualization literature surrounding approaches to recording and reporting, and from literature promoting reflection in research and in practice.

The proper recording and reporting of research process and findings is necessary for establishing reliability and transferability in design study. Despite this, existing visualization methodology offers little guidance on how to do this effectively, and existing tools offer little support for incorporating these practices into the design study process. Most closely related to our notion of recording and reporting is the large body of research in visualization and data analysis focused on analytic provenance. Analytic provenance is broadly concerned with capturing the history of changes throughout the analysis pipeline [136]. The provenance literature covers theoretical approaches to characterizing provenance [136],[236],[237] as well as mechanisms to facilitate provenance through the various stages of analysis [238]–[241].

Based on a comprehensive review of this literature, Ragan *et al.* [136] offer an organizational framework that characterizes analytic provenance by type and by purpose. *Data provenance* traces the changes and movement of data. *Visualization provenance* captures graphical views and visualization states during the analysis of data using a visual analytics system, and *interaction provenance* captures the actions taken to generate those states. *Insight provenance* captures insights gained by users during this analysis, primarily through annotation. Finally, *rational provenance* traces the reasoning and intentions behind interactions and decisions during analysis. Purposes for analytic provenance include recall, replication, action recovery, collaborative communication, presentation, and meta-analysis. We have yet to see the notion of analytic provenance applied to capturing the evolution of the design process and the emergence of research insight through the use of a visualization system. We see potential value in viewing recording and reporting of design study artifacts through the lens of this organizational framework.

Although recording and reporting are necessary for capturing and communicating artifacts, reflection is critical for the *analysis* of artifacts. Reflection supports leveraging artifacts to inform design, and it is also a key mechanism for contextualizing artifacts with the necessary detail to promote transferability and reliability. Our understanding of reflection draws from design study methodology and from the literature promoting reflection in research and in practice.

Like transferability, reflection is cited in design study methodology as another core component to producing actionable contributions and moving design insights toward broader visualization knowledge [8], [26]. In particular, reflecting on the design study process, findings, and lessons learned can lead to the confirmation, refinement, or rejection of existing guidelines, or to the proposal of new guidelines. Therefore, this kind of reflection is defined as the third type of design study contribution and is identified as one of the final stages of the nine-stage methodological framework [8]. The practice of reflecting on methodological approaches to design study has been adopted by researchers, and it is now common to find this third type of contribution in design study publications [26]. Still, guidance on how to reflect, as well as mechanisms to support reflection in design study, is lacking. Fortunately, ample guidance on reflection is offered in disciplines spanning both research and practice.

In investigating our strategy, we turn to the body of literature on *reflective practice* [242]. Reflective practice describes "the process by which professionals become aware of their implicit knowledge and learn from their experience" [242]. This approach to turning experience into learning builds on the established notion that tacit knowledge is acquired through experience and cannot be expressed in clear rules [243], [244]. Reflective practice revolves around two main activities. Reflection-*in*-action involves thinking on your feet during an intervention. Reflection-*on*-action involves reflecting postintervention to inform future practice. Extending this approach, *critically reflective practice* introduces a third activity, reflection-*for*-action, which involves reflecting in preparation for an intervention. Critically reflective practice additionally encourages practitioners to reflect on the power of thought in informing practice, the role of emotion in shaping practice, and the moral-political factors and their implications for practice [245]. The goal of critically reflective practice is to synthesize experience, reflection, self-awareness, and critical thinking in order to modify approaches to practice [11], [245].

Both reflective practice and critically reflective practice have been highly influential and have been embraced by many disciplines, most openly by nursing and healthcare professionals [246] and by the design community [243]. As its name implies, reflective practice is geared toward the work of professional practitioners who repeatedly encounter the same set of situations [242]. This work is fundamentally different from that of applied researchers, who use practice to identify and contribute to open research questions. To address this difference, there have been efforts within the design research community to incorporate reflective practice into research methodology. In information systems research, the action design research method [5] defines *reflection and learning* as a stage of research, occurring continuously through the exploratory and development stages, and critical to generating design knowledge through practice. Additionally, in participatory design, a proposed conceptual framework for reflection [247] guides researchers and practitioners in critically reflecting on tacit qualities of participatory design through the lenses of epistemology, values, stakeholders, and outcomes. The framework is designed to support researchers in both assessing and communicating the rigor and accountability of their research.

7.3.3 Important Areas for Future Visualization Research

We conclude our investigation into recording, reflecting, and reporting on artifacts in design study with a discussion of important areas for future visualization research. Our exploration of artifacts in design study points to two important areas for future applied visualization research: characterizing the span of artifacts generated in design study and evaluating the role of artifacts in increasing the reliability and transferability of the research process and findings. We discuss each of these in turn and suggest possible research strategies.

Before we can use artifacts to promote the reliability and transferability of our research, we need to establish what artifacts look like in the context of design study. More specifically, we need to characterize the range of artifacts generated throughout the design process and the span of learning that these artifacts externalize. Based on our preliminary research, we suggest starting with the systematic, qualitative analysis of example artifacts found within the visualization literature — published as supplemental material, reported as contributions in design study publications, and including examples of field notes, thick description, and audit trails. Conducting a round of qualitative coding [248] on this set of example artifacts can help to identify initial themes and classifications. Published artifacts, however, likely represent only a small fraction of the artifacts generated throughout the design process. In addition, we as readers can only speculate, at best, about the learning that these artifacts externalize. Supplementing the analysis of published artifacts with interviews with publication authors can reveal a broader range of artifacts and can help to clarify the associated learning. Recording, transcribing (or reflectively transcribing, as described in section 5.4.1), and qualitatively analyzing these interviews can help to identify additional themes and classifications.

In addition to drawing from the visualization literature, workshops with fellow design study researchers provide another opportunity to gather examples of artifacts and to collectively reflect on the associated learning. Such workshops can also be used to reflect on the kinds of learning that are not currently being captured in artifacts, but could yield valuable visualization knowledge. We propose extending any characterization of existing design study artifacts to include other forms of artifacts (e.g., adapted from other fields) that may help to externalize this broader span of learning. The resulting characterization would provide a starting point, to be further developed and refined by the applied visualization research community.

Once we have developed a better understanding of design study artifacts, the next step is to investigate the role of artifacts in increasing the reliability and transferability of the research process and findings. One approach would be to base the investigation in a series of interviews with applied visualization researchers spanning different research styles within the visualization community. Prior to interviews, participants could be asked to review a set of curated artifacts alongside a publication, and to think deeply about what elements of the artifacts are important for increasing the reliability and/or transferability of the research and how these elements could be strengthened. Interviews can be recorded and reflectively transcribed, and qualitative coding can be used to identify themes and patterns. As a preliminary step, e.g., as part of the introduction to the study, participants could be interviewed about their current approaches to evaluating the reliability and transferability of applied visualization research publications. Anticipated results would be a set of guidelines for using artifacts to increase reliability and transferability of applied visualization.

As with the characterization, these guidelines would provide a starting point, to be expanded and refined by other researchers within the visualization community, based on their own research processes and the range of artifacts that they generate. In addition to offering guidance, this work would provide evidence that artifacts can in fact increase reliability and transferability, and would convince visualization researchers that the additional effort required to curate and publish artifacts is in fact worthwhile.

Our exploration into recording, reflecting, and reporting in design study points to three important areas for future visualization research: characterizing reliability and transferability in design study research; characterizing the role of recording, reflecting, and reporting on artifacts in increasing reliability and transferability; and designing methods and mechanisms to support these activities throughout the design process. Fortunately, recent theoretical work within the visualization community seeks to define concepts of reliability, transferability, and rigor more broadly in the context of applied visualization research [212]. An important next step would therefore be to re-evaluate and evolve our work, and our proposed strategy, in light of these definitions.

CHAPTER 8

CONCLUSION

This dissertation explores the contextual, human-centered nature of applied visualization research and the degree to which action design research can help to address existing threats to the reliability of the research process and resulting technology artifacts. Our exploration is grounded in a formative design study with poetry scholars, which revealed gaps in existing visualization methodology and motivated us to look to fields beyond visualization for guidance on conducting research in highly collaborative, contextual, and dynamic environments. Our search led us to the field information systems, and to the method of action design research (ADR) as a potentially useful model for providing this guidance. By incorporating established social science methods, ADR provides a framework for navigating and leveraging the role of people, context, and disruption in shaping the design process.

Through preliminary theoretical research into ADR and its relationship to existing visualization models, we obtained initial validation of the framework and demonstrated its applicability to visualization design research. Over the course of two consecutive design studies, we then tested the framework in the wild and experimented with ways of supporting the framework throughout the design process. In the first design study with global health experts, certain elements of the framework proved particularly useful for supporting our research, and certain strategies emerged for supporting and enhancing these elements throughout the design process. In the second design study with astronomers and astrophysicists, we further tested and validated these results and further investigated and developed these strategies.

Our exploratory results validate ADR as a useful model for strengthening elements of applied visualization research. At the broadest level, ADR strengthens applied visualization research by conditioning researchers to attend to and leverage the role of people and context in shaping the design process and results. On a practical level, ADR offers a valuable framework for conducting applied visualization research by providing structure and guidance for facilitating, enhancing, and formalizing the learning that occurs throughout the design process.

Our exploratory results also reveal significant gaps in ADR for applied visualization research. Namely, ADR fails to adequately support reliability and transferability in applied visualization research, and would benefit from a broader notion of artifacts and mechanisms to support the framework throughout the design process. As part of a reflective synthesis of our exploratory results, we speculate on a way to address these gaps through recording, reflecting, and reporting on artifacts in design study. We present this potential strategy as an important area for future visualization research.

The primary contributions of this dissertation include an articulation of the gaps in existing visualization methodology, grounded in our formative design study; an exploration of action design research for applied visualization design, grounded in theory and applied in practice; and a reflective synthesis of the exploratory results. Secondary contributions stem from the results of the three design studies and reflect the development of our research thinking around artifacts, the emergent design process, and the generation of visualization knowledge.

REFERENCES

[1] N. McCurdy, J. Lein, K. Coles, and M. Meyer, "Poemage: Visualizing the sonic topology of a poem," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 439–448, Jan. 2016.

[2] N. McCurdy, J. Dykes, and M. Meyer, "Action design research and visualization design," in *Proc. 6th Workshop. Beyond Time Errors: Novel Evaluation Methods Inf. Vis.*, Oct. 2016, pp. 10–18, doi:10.1145/2 993 901.2 993 916.

[3] N. McCurdy, J. Gerdes, and M. Meyer, "A framework for externalizing implicit error using visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 925–935, Jan. 2019.

[4] N. McCurdy and M. Meyer, "Galstamps: Analyzing real and simulated galaxy observations," in *Proc. IEEE Vis Conf.*, Oct. 2019, pp. 276–280.

[5] M. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, "Action design research," *Manage. Inf. Syst. Quart.*, vol. 35, no. 1, pp. 37–56, Mar. 2011.

[6] R. Buchanan, "Wicked problems in design thinking," *Des. Issues*, vol. 8, no. 2, pp. 5–21, Apr. 1992.

[7] S. McKenna, D. Mazur, J. Agutter, and M. Meyer, "Design activity framework for visualization design," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2191–2200, Dec. 2014.

[8] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2431–2440, Dec. 2012.

[9] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner, "The nested blocks and guidelines model," *Inf. Vis.*, vol. 14, no. 3, pp. 234–249, Jul. 2015.

[10] T. Munzner, "A nested model for visualization design and validation," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 921–928, Nov.-Dec. 2009.

[11] E. Kerzner, S. Goodwin, J. Dykes, S. Jones, and M. Meyer, "A framework for creative visualization-opportunities workshops," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 748–758, Jan. 2018.

[12] J. Wood, A. Kachkaev, and J. Dykes, "Design exposition with literate visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 759–768, Jan. 2018.

[13] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2431–2440, Dec. 2012.

[14] G. R. Hayes, "The relationship of action research to human-computer interaction," *ACM Trans. Comput.-Human Interaction*, vol. 18, no. 3, p. 15, Jul. 2011.

[15] D. J. Greenwood and M. Levin, *Introduction to Action Research: Social Research for Social Change*. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2006.

[16] J. McNiff, *Action Research: Principles and Practice*. New York, NY, USA: Routledge, 2013.

[17] N. McCurdy, V. Srikumar, and M. Meyer, "Rhymedesign: A tool for analyzing sonic devices in poetry," in *Proc. 4th Workshop. Comput. Linguistics Literature*, Jun. 2015, pp. 12–22.

[18] A. Forbes, "Highlights from the IEEE VIS 2015 Arts Program," *Leonardo*, vol. 51, no. 1, pp. 70–70, Feb. 2018.

[19] J. G. Lein, N. McCurdy, and A. Hurtado, "Deep in poetry: Visualizing texts' sonic depths in 3d," *Leonardo*, vol. 51, no. 1, pp. 80–82, Feb. 2017.

[20] N. K. Denzin and Y. S. Lincoln, *Handbook of Qualitative Research.* Thousand Oaks, CA, USA: SAGE Publications, Inc., 1994.

[21] Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*. Thousand Oaks, CA, USA: SAGE Publications, Inc., 1985, vol. 75.

[22] R. H. von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *Manage. Inf. Syst. Quart.*, vol. 28, no. 1, pp. 75–105, Mar. 2004.

[23] E. T. Stringer, *Action Research*. Thousand Oaks, CA, USA: SAGE Publications, Inc., Jun. 2013.

[24] H. W. Rittel and M. M. Webber, "Dilemmas in a general theory of planning," *Policy Sciences*, vol. 4, no. 2, pp. 155–169, June 1973.

[25] I. Dillingham, J. Dykes, and J. Wood, "A design, analysis and evaluation model to support the visualization designer-user," in *Poster presented at IEEE Conf. Inf. Vis.*, Oct. 2012.

[26] M. Sedlmair, "Design study contributions come in different guises: Seven guiding scenarios," in *Proc. 6th Workshop. Beyond Time Errors: Novel Evaluation Methods Inf. Vis.*, Oct. 2016, pp. 152–161.

[27] C. J. Zimmerman, H. T. Wessels, and R. Vatrapu, "Building a social newsroom: Visual analytics for social business intelligence," in *2015 IEEE 19th Int. Enterprise Distributed Object Comp. Workshop*, Oct. 2015, pp. 160–163.

[28] S. T. March and G. F. Smith, "Design and natural science research on information technology," *Decis. Support Syst.*, vol. 15, no. 4, pp. 251–266, Dec. 1995.

[29] W. J. Orlikowski and C. S. Iacono, "Research commentary: Desperately seeking the "it" in it research—a call to theorizing the it artifact," *Inf. Syst. Res.*, vol. 12, no. 2, pp. 121–134, Jun. 2001.

[30] S. Purao, O. Henfridsson, M. Rossi, and M. Sein, "Ensemble artifacts: From viewing to designing in action design research," *Syst., Signs & Actions*, vol. 7, no. 1, pp. 73–81, Jan. 2013.

[31] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen *et al.*, "Technology probes: Inspiring design for and with families," in *Proc. ACM Conf. Human Factors Comput. Syst.*, Apr. 2003, pp. 17–24.

[32] D. Lloyd and J. Dykes, "Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2498–2507, Dec. 2011.

[33] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale, "Grounded evaluation of information visualizations," in *Proc. 2008 Workshop Beyond Time Errors: Novel Evaluation Methods Inf. Vis.*, Apr. 2008, p. 6.

[34] K. Andrews, "Evaluation comes in many guises," in *Proc. 2008 Workshop Beyond time and errors Position Paper*, Apr. 2008, pp. 7–8.

[35] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Seven guiding scenarios for information visualization evaluation," University of Calgary, Calgary, Alberta, Canada, Tech. Rep. 2011-992-04, Jan. 2011.

[36] B. Shneiderman and C. Plaisant, "Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies," in *Proc. 2006 AVI Workshop. Beyond Time Errors: Novel Evaluation Methods Inf. Vis.*, May 2006, pp. 1–7.

[37] S. Goodwin, J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood, "Creative user-centered visualization design for energy analysts and modelers," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2516–2525, Oct. 2013.

[38] M. Meyer, T. Munzner, and H. Pfister, "Mizbee: A multiscale synteny browser," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 897–904, Oct. 2009.

[39] S. K. Card and J. Mackinlay, "The structure of the information visualization design space," in *Proc. VIZ'97: IEEE Vis Conf. Inf. Vis Symp Parallel Rendering Symp.*, Oct. 1997, pp. 92–99.

[40] E. H.-h. Chi and J. T. Riedl, "An operator interaction framework for visualization systems," in *Proc. IEEE Symp. Inf. Vis.*, Oct. 1998, pp. 63–70.

[41] S. Carpendale, "Evaluating information visualizations," in *Inf. Vis. Lecture Notes in Comput. Sci.*, vol. 4950. Berlin, Heidelberg: Springer, 2008, pp. 19–45.

[42] M. Tory and T. Moller, "Human factors in visualization research," *IEEE Trans. Vis. Comput. Graphics*, vol. 10, no. 1, pp. 72–84, Jun. 2004.

[43] J. J. Van Wijk, "The value of visualization," in *VIS 05. IEEE Vis.*, Oct. 2005, pp. 79–86.

[44] M. Meyer and J. Dykes, "Reflection on reflection in applied visualization research," *IEEE Comput. Graphics Appl.*, vol. 38, no. 6, pp. 9–16, Nov. 2018.

[45] K. Lewin, "Action research and minority problems," *J. Social Issues*, vol. 2, no. 4, pp. 34–46, Nov. 1946.

[46] E. Lieberman, J.-B. Michel, J. Jackson, T. Tang, and M. A. Nowak, "Quantifying the evolutionary dynamics of language," *Nature*, vol. 449, no. 7163, pp. 713–716, Oct. 2007.

[47] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1137–1144, Oct. 2009.

[48] F. Moretti, *Distant Reading*. Brooklyn, New York, NY. USA: Verso Books, Jun. 2013.

[49] P. DeCamp, A. Frid-Jimenez, J. Guiness, and D. Roy, "Gist icons: Seeing meaning in large bodies of literature," in *Proc. IEEE Symp. Inf. Vis.*, Oct. 2005.

[50] C. Collins, S. Carpendale, and G. Penn, "Docuburst: Visualizing document content using language structure," in *Comput. Graphics Forum*, vol. 28, no. 3. Wiley Online Library, Jun. 2009, pp. 1039–1046.

[51] J.-D. Fekete and N. Dufournaud, "Compus: Visualization and analysis of structured documents for understanding social life in the 16th century," in *Proc. 5th ACM Conf. Digit. Libraries*. ACM, Jun. 2000, pp. 47–55.

[52] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow, "Visualizing the non-visual: Spatial analysis and interaction with information from text documents," in *Proc. IEEE 1995 Vis. Conf.*, Oct. 1995, pp. 51–58.

[53] W. B. Paley, "Textarc: Showing word frequency and distribution in text," in *Proc. IEEE Symp. Inf. Vis.*, vol. 2002, Oct. 2002.

[54] C. Collins, F. B. Viegas, and M. Wattenberg, "Parallel tag clouds to explore and analyze faceted text corpora," in *2009 IEEE Symp. Vis. Analytics Sci. Technol.*, Oct. 2009, pp. 91–98.

[55] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: Supporting investigative analysis through interactive visualization," *Inf. Vis.*, vol. 7, no. 2, pp. 118–132, Jun. 2008.

[56] M. Correll, M. Witmore, and M. Gleicher, "Exploring collections of tagged text for literary scholarship," in *Comput. Graph. Forum*, vol. 30, no. 3. Wiley Online Library, Jun. 2011, pp. 731–740.

[57] A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant, "Discovering interesting usage patterns in text collections: Integrating text mining with visualization," in *Proc. 16th ACM Conf. on Conf. Inf. Knowl. Manage.*, Nov. 2007, pp. 213–222.

[58] F. Van Ham, M. Wattenberg, and F. B. Viégas, "Mapping text with phrase nets," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1169–1176, Oct. 2009.

[59] M. Wattenberg, "Arc diagrams: Visualizing structure in strings," in *2002 IEEE Symp. Inf. Vis.*, Oct. 2002, pp. 110–116.

[60] M. Wattenberg and F. B. Viégas, "The word tree, an interactive visual concordance," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1221–1228, Oct. 2008.

[61] A. Muralidharan and M. A. Hearst, "Supporting exploratory text analysis in literature study," *Lit. Ling. Comput.*, vol. 28, no. 2, pp. 283–295, Dec. 2013.

[62] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen, "Rule-based visual mappings–with a case study on poetry visualization," in *Comput. Graph. Forum*, vol. 32, no. 3pt4, Jun. 2013, pp. 381–390.

[63] M. Chaturvedi, G. Gannod, L. Mandell, H. Armstrong, and E. Hodgson, "Myopia: A visualization tool in support of close reading," *Digit. Humanities*, Jul. 2012.

[64] T. Clement, D. Tcheng, L. Auvil, B. Capitanu, and M. Monroe, "Sounding for meaning: Using theories of knowledge representation to analyze aural patterns in texts," *Digit. Humanities Quart.*, vol. 7, Jan. 2013.

[65] T. Clement, "Distant listening or playing visualizations pleasantly with the eyes and ears." *Digital Studies/Le champ numérique*, vol. 3, no. 2, Jul. 2012.

[66] S. Fish. (2012, Jan. 23) Mind your ps and bs: the digital humanities and interpretation. [Online]. Available: https://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation/

[67] S. Ramsay, *Reading Machines: Toward an Algorithmic Criticism*. Urbana, IL, USA: Univ. of Illinois Press, Dec. 2011.

[68] C. Reas and B. Fry, "Processing. org: A networked context for learning computer programming," in *ACM SIGGRAPH 2005 web program*, Jul. 2005, pp. 14–es.

[69] J. Lein, "Sounding the surfaces: Computers, context, and poetic consequence," *Western Humanities Rev.*, pp. 84–109, Fall 2014.

[70] gicentre utilities. apr. 14, 2014. [Online]. Available: http://www.gicentre.net/utils/ [Accessed: Nov. 15, 2015]

[71] B. Alper, N. H. Riche, G. Ramos, and M. Czerwinski, "Design study of linesets, a novel set visualization technique," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2259–2267, Nov. 2011.

[72] J. Ellson, E. Gansner, L. Koutsofios, S. C. North, and G. Woodhull, "Graphviz—open source graph drawing tools," in *Int. Symp. Graph Drawing*. Springer, Sep. 2001, pp. 483–484.

[73] M. Bastian, S. Heymann, M. Jacomy *et al.*, "Gephi: An open source software for exploring and manipulating networks." *3rd Int. AAAI Conf. Weblogs Social Media*, vol. 8, pp. 361–362, Mar. 2009.

[74] K. Dinkla, M. J. van Kreveld, B. Speckmann, and M. A. Westenberg, "Kelp diagrams: Point set membership visualization," in *Comput. Graph. Forum*, vol. 31, no. 3pt1, Jun. 2012, pp. 875–884.

[75] W. Meulemans, N. H. Riche, B. Speckmann, B. Alper, and T. Dwyer, "Kelpfusion: A hybrid set visualization technique," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 11, pp. 1846–1858, Apr. 2013.

[76] C. Collins, G. Penn, and S. Carpendale, "Bubble sets: Revealing set relations with isocontours over existing visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1009–1016, Oct. 2009.

[77] H. Byelas and A. Telea, "Visualizing metrics on areas of interest in software architecture diagrams," in *IEEE Pacific Vis. Symp.*, Apr. 2009, pp. 33–40.

[78] J. Heer and D. Boyd, "Vizster: Visualizing online social networks," in *IEEE Symp. Inf. Vis.*, Oct. 2005, pp. 32–39.

[79] S.-J. Luo, C.-L. Liu, B.-Y. Chen, and K.-L. Ma, "Ambiguity-free edge-bundling for interactive graph visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 5, pp. 810–821, Jun. 2012.

[80] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li, "Geometry-based edge clustering for graph visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1277–1284, Oct. 2008.

[81] O. Ersoy, C. Hurter, F. V. Paulovich, G. Cantareiro, and A. Telea, "Skeleton-based edge bundling for graph visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2364–2373, Nov. 2011.

[82] A. Lambert, R. Bourqui, and D. Auber, "Winding roads: Routing edges into bundles," in *Proc. Comput. Graphics Forum*, vol. 29, no. 3, Jun. 2010, pp. 853–862.

[83] S. Pupyrev, L. Nachmanson, S. Bereg, and A. E. Holroyd, "Edge routing with ordered bundles," in *Int. Symp. Graph Drawing*, Sep. 2012, pp. 136–147.

[84] M. Nöllenburg, "An improved algorithm for the metro-line crossing minimization problem," in *Int. Symp. Graph Drawing*, Sep. 2010, pp. 381–392.

[85] M. Suri. (2015, Mar. 13) Don't expect math to make sense. [Online]. Available: https://www.nytimes.com/2015/03/14/opinion/manil-suri-on-pi-day-celebrate-maths-enigmas.html

[86] K. Coles and J. Lein, "Turbulence and temporality: (re)visualizing poetic time. things my computer taught me about poems." in *Proc. 2014 Modern Lang. Assoc.*, Chicago, IL, Jan. 2014.

[87] K. Coles, "Slippage, spillage, pillage, bliss: Close reading, uncertainty, and machines," *Western Humanities Rev.*, pp. 39–65, Fall 2014.

[88] J. McGann and L. Samuels, "Deformance and interpretation," *New Literary Hist.*, vol. 30, no. 1, pp. 25–56, Jan. 1999.

[89] E. Irizarry, "Tampering with the text to increase awareness of poetry's art.(theory and practice with a hispanic perspective)," *Literary Linguistic Comput.*, vol. 11, no. 4, pp. 155–162, Dec. 1996.

[90] S. Ramsay, "The hermeneutics of screwing around; or what you do with a million books," in *Proc. Playing Technol. Hist. Conf.*, vol. 17, Apr. 2010, p. 103.

[91] W. S. Cleveland, "A model for studying display methods of statistical graphics," *J. Comput. Graphical Statist.*, vol. 2, no. 4, pp. 323–343, Dec. 1993.

[92] T. Munzner, "Process and pitfalls in writing information visualization research papers," in *Inf. Vis. : Human-Centered Issues and Perspectives*, Mar. 2008, pp. 134–153.

[93] L. N. Carroll, A. P. Au, L. T. Detwiler, T.-c. Fu, I. S. Painter, and N. F. Abernethy, "Visualization and analytics tools for infectious disease epidemiology: A systematic review," *J. Biomed. Inf.*, vol. 51, pp. 287–298, Oct. 2014.

[94] A. Ramanathan, L. L. Pullum, C. A. Steed, S. Quinn, C. S. Chennubhotla, and T. L. Parker, "Integrating heterogeneous healthcare datasets and visual analytics for disease biosurveillance and dynamics," in *Proc. 3rd IEEE Workshop Vis. Text Analytics*, Oct. 2013.

[95] S. Afzal, R. Maciejewski, and D. S. Ebert, “Visual analytics decision support environment for epidemic modeling and response evaluation,” in *Proc. IEEE Conf. Vis. Analytics Sci. Technol.*, Oct. 2011, pp. 191–200.

[96] C. Hayes, M. Jilani, A. Yañez, M. Connolly, and J. Duggan, “Pandemcap: Decision support tool for epidemic management,” in *Proc. 2017 IEEE Workshop Vis. Analytics Healthcare*, Oct. 2017.

[97] Y. Livnat, T.-M. Rhyne, and M. Samore, “Epinome: A visual-analytics workbench for epidemiology data,” *IEEE Comput. Graphics Appl.*, vol. 32, no. 2, pp. 89–95, Feb. 2012.

[98] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders, “A pandemic influenza modeling and visualization tool,” *J. Vis. Lang. & Comput.*, vol. 22, no. 4, pp. 268–278, Aug. 2011.

[99] W. Van den Broeck, C. Gioannini, B. Gonçalves, M. Quaggiotto, V. Colizza, and A. Vespignani, “The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale,” *BMC Infectious Diseases*, vol. 11, no. 1, p. 37, Dec. 2011.

[100] J. W. Loonsk, “Biosense: Implementation of a national early event detection and situational awareness system,” *Morb. Mortal. Wkly. Rep.*, pp. 53–55, Sep. 2004.

[101] P. H. Gesteland, Y. Livnat, N. Galli, M. H. Samore, and A. V. Gundlapalli, “The epicanvas infectious disease weather map: An interactive visual exploration of temporal and spatial correlations,” *J. Amer. Med. Inf. Assoc.*, vol. 19, no. 6, pp. 954–959, Nov. 2012.

[102] K. Brown, J. Pavlin, J. Mansfield, E. Elbert, V. Foster, and P. Kelley, “Identification and investigation of disease outbreaks by essence,” *J. Urban Health*, vol. 80, pp. i119–i119, Mar. 2003.

[103] J. U. Espino, M. Wagner, C. Szczepaniak, F. Tsui, H. Su, R. Olszewski, Z. Liu, W. Chapman, X. Zeng, L. Ma *et al.*, “Removing a barrier to computer-based outbreak and disease surveillance—the rods open source project,” *Morb. Mortal. Wkly. Rep.*, pp. 32–39, Sep. 2004.

[104] B. Y. Reis, C. Kirby, L. E. Hadden, K. Olson, A. J. McMurry, J. B. Daniel, and K. D. Mandl, “Aegis: A robust and scalable real-time public health surveillance system,” *J. Amer. Med. Inf. Assoc.*, vol. 14, no. 5, pp. 581–588, Sep. 2007.

[105] C. C. Freifeld, K. D. Mandl, B. Y. Reis, and J. S. Brownstein, “Healthmap: Global infectious disease monitoring through automated classification and visualization of internet media reports,” *J. Amer. Med. Inf. Assoc.*, vol. 15, no. 2, pp. 150–157, Mar. 2008.

[106] M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” *Human Factors*, vol. 37, no. 1, pp. 32–64, Mar. 1995.

[107] Y. Livnat, J. Agutter, S. Moon, and S. Foresti, “Visual correlation for situational awareness,” in *IEEE Symp. Inf. Vis.*, Oct. 2005, pp. 95–102.

[108] S. Declich and A. O. Carter, “Public health surveillance: Historical origins, methods and evaluation.” *Bulletin World Health Org.*, vol. 72, no. 2, p. 285, Feb. 1994.

[109] D. G. Koutsonanos, “Public health surveillance systems for disease monitoring, situational awareness, and decision making support,” Ph.D. dissertation, Dept. Microbiology Immunology. Emory University, Atlanta, GA, USA, 2014.

[110] M. S. Louis, “Global health surveillance,” *Morb. Mortal. Wkly. Rep. Surveill. Summ.*, vol. 61, no. Suppl., pp. 15–19, Jul. 2012.

[111] T. G. Savel, S. Foldy, C. for Disease Control, Prevention *et al.*, “The role of public health informatics in enhancing public health surveillance,” *Morb. Mortal. Wkly. Rep. Surveill Summ*, vol. 61, no. Suppl, pp. 20–4, Jul. 2012.

[112] G. C. Bowker and S. L. Star, *Sorting Things Out: Classification and its Consequences*. Cambridge, MA, USA: MIT Press, 2000.

[113] M. C. Hoyos, R. S. Morales, and R. Akhavan-Tabatabaei, “Or models with stochastic components in disaster operations management: A literature survey,” *Comput. Ind. Eng.*, vol. 82, pp. 183–197, Apr. 2015.

[114] R. W. Perry and M. K. Lindell, “Preparedness for emergency response: Guidelines for the emergency planning process,” *Disasters*, vol. 27, no. 4, pp. 336–350, Dec. 2003.

[115] J. Rasmussen, “Risk management in a dynamic society: A modelling problem,” *Saf. Sci.*, vol. 27, no. 2-3, pp. 183–213, November-December 1997.

[116] P. Carayon and K. E. Wood, “Patient safety: The role of human factors and systems engineering,” *Stud. Health Technol. Inform.*, vol. 153, pp. 23–46, Mar. 2010.

[117] World Health Organization. (2016, Apr.) Global health observatory. [Online]. Available: http://www.who.int/gho/tb/en [Accessed: Feb. 10, 2017]

[118] B. C. Choi, “The past, present, and future of public health surveillance,” *Scientifica*, vol. 2012, Oct. 2012.

[119] World Health Organization, *Data quality review: Module 1: Framework and metrics*, 2017. [Online]. Available: https://apps.who.int/iris/handle/10665/259224

[120] S. B. Thacker, J. R. Qualters, L. M. Lee, and Centers for Disease Control and Prevention, “Public health surveillance in the united states: Evolution and challenges,” *Morb. Mortal. Wkly. Rep. Surveill. Summ.*, vol. 61, no. Suppl, pp. 3–9, Jul. 2012.

[121] J. S. Brown and P. Duguid, *The Social Life of Information: Updated, With a New Preface*. Brighton, MA, U.S.A.: Harvard Business Review, Feb. 2017.

[122] J. Hullman, M. Kay, Y.-S. Kim, and S. Shrestha, “Imagining replications: Graphical prediction & discrete visualizations improve recall & estimation of effect uncertainty,” *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 446–456, Aug. 2018.

[123] N. Kijmongkolchai, A. Abdul-Rahman, and M. Chen, “Empirically measuring soft knowledge in visualization,” in *Comput. Graph. Forum*, vol. 36, no. 3, Jun. 2017, pp. 73–85.

[124] Y.-S. Kim, K. Reinecke, and J. Hullman, “Explaining the gap: Visualizing one’s predictions improves recall and comprehension of data,” in *Proc. 2017 CHI Conf. Human Factors Comput. Syst.*, May 2017, pp. 1375–1386.

[125] Y.-S. Kim, K. Reinecke, and J. Hullman, "Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 760–769, Aug. 2018.

[126] P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner, "The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics," in *Proc. IEEE Conf. Vis. Analytics Sci. Technol.*, Oct. 2017.

[127] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver, "Data, information, and knowledge in visualization," *IEEE Comput. Graphics Appl.*, vol. 29, no. 1, Dec. 2009.

[128] X. Wang, D. H. Jeong, W. Dou, S.-w. Lee, W. Ribarsky, and R. Chang, "Defining and applying knowledge conversion processes to a visual analytics system," *Comput. Graphics*, vol. 33, no. 5, pp. 616–623, Oct. 2009.

[129] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz, "Overview and state-of-the-art of uncertainty visualization," in *Sci. Vis. Mathematics Vis.*, Sep. 2014, pp. 3–27.

[130] N. Gershon, "Visualization of an imperfect world," *IEEE Comput. Graphics Appl.*, vol. 18, no. 4, pp. 43–45, Jul. 1998.

[131] C. Kinkeldey, A. M. MacEachren, and J. Schiewe, "How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies," *Cartographic J.*, vol. 51, no. 4, pp. 372–386, Nov. 2014.

[132] K. Potter, P. Rosen, and C. R. Johnson, "From quantification to visualization: A taxonomy of uncertainty visualization approaches," in *IFIP Working Conf. Uncertainty Quant. Sci. Comput.*, Aug. 2012, pp. 226–249.

[133] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, "A typology for visualizing uncertainty," in *Vis. Data Anal.*, vol. 5669, Mar. 2005, pp. 146–158.

[134] L. D. Edwards and E. S. Nelson, "Visualizing data certainty: A case study using graduated circle maps," *Cartographic Perspectives*, no. 38, pp. 19–36, Mar. 2001.

[135] P. Chen, B. Plale, Y.-W. Cheah, D. Ghoshal, S. Jensen, and Y. Luo, "Visualization of network data provenance," in *Proc. 19th Int. Conf. High Perform. Comput.*, Dec. 2012, pp. 1–9.

[136] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen, "Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 31–40, Aug. 2016.

[137] K. Xu, S. Attfield, T. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj, "Analytic provenance for sensemaking: A research agenda," *IEEE Comput. Graphics Appl.*, vol. 35, no. 3, pp. 56–64, May 2015.

[138] R. Walker, A. Slingsby, J. Dykes, K. Xu, J. Wood, P. H. Nguyen, D. Stephens, B. W. Wong, and Y. Zheng, "An extensible framework for provenance in human terrain visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2139–2148, Oct. 2013.

[139] S. B. Thacker, R. G. Parrish, and F. L. Trowbridge, “A method for evaluating systems of epidemiological surveillance,” *World Health Statist. Quart.*, vol. 41, no. 1, Jan. 1988.

[140] S. Brookfield, “Critically reflective practice,” *J. Continuing Ed. Health Professions*, vol. 18, no. 4, pp. 197–205, Sep. 1998.

[141] S. Thompson and N. Thompson, *The Critically Reflective Practioner*. New York, NY, USA: Palgrave Macmillan, 2008.

[142] I. Nonaka and H. Takeuchi, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. London, UK: Oxford University Press, 1995.

[143] K. G. Villholth, E. Lopez-Gunn, K. Conti, A. Garrido, and J. Van Der Gun, *Advances in Groundwater Governance*. Boca Raton, FL, U.S.A: CRC Press, 2017.

[144] E. Trist, “The evolution of socio-technical systems,” *Occasional Paper / Ontario Quality of Working Life Centre*, no. 2, Jun. 1981.

[145] C. Pappas and I. Williams, “Grey literature: Its emerging importance,” *J. Hospital Librarianship*, vol. 11, no. 3, pp. 228–234, Jul. 2011.

[146] M. Alur, “Forgotten millions: A case of cultural and systemic bias,” *Support Learn.*, vol. 22, no. 4, pp. 174–180, Nov. 2007.

[147] N. Boukhelifa, M.-E. Perrin, S. Huron, and J. Eagan, “How data workers cope with uncertainty: A task characterisation study,” in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2017, pp. 3645–3656.

[148] C. C. Freifeld, R. Chunara, S. R. Mekaru, E. H. Chan, T. Kass-Hout, A. A. Iacucci, and J. S. Brownstein, “Participatory epidemiology: Use of mobile phones for community-based health reporting,” *PLoS Medicine*, vol. 7, no. 12, p. e1000376, Dec. 2010.

[149] J. A. Sacks, E. Zehe, C. Redick, A. Bah, K. Cowger, M. Camara, A. Diallo, A. N. I. Gigo, R. S. Dhillon, and A. Liu, “Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea,” *Global Health: Sci. Pract.*, vol. 3, no. 4, pp. 646–659, Dec. 2015.

[150] T. L. Lash, M. P. Fox, and A. K. Fink, *Applying Quantitative Bias Analysis to Epidemiologic Data*. Berlin, Germany: Springer Science & Business Media, 2011.

[151] J. Rowley, “The wisdom hierarchy: Representations of the dikw hierarchy,” *J. Inf. Sci.*, vol. 33, no. 2, pp. 163–180, Apr. 2007.

[152] T. H. Davenport and L. Prusak, *Working Knowledge: How Organizations Manage What They Know*. Brighton, MA, U.S.A.: Harvard Business Press, 1998.

[153] D. Clark, “Understanding and performance.” [Online]. Available: http://www.nwlink.com/ donclark/performance/understanding.html [Accessed: Jan. 10, 2018]

[154] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe, “Chartaccent: Annotation for data-driven storytelling,” in *Proc. IEEE Pacific Vis. Symp.*, Apr. 2017, pp. 230–239.

[155] A. Satyanarayan and J. Heer, "Authoring narrative visualizations with ellipsis," in *Proc. Comput. Graphics Forum*, vol. 33, no. 3, Jun. 2014, pp. 361–370.

[156] J. Zhao, M. Glueck, S. Breslav, F. Chevalier, and A. Khan, "Annotation graphs: A graph-based visualization for meta-analysis of data based on user-authored annotations," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 261–270, Aug. 2017.

[157] J. Hullman, E. Adar, and P. Shah, "The impact of social information on visual judgments," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, May 2011, pp. 1461–1470.

[158] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon, "Manyeyes: A site for visualization at internet scale," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, Nov. 2007.

[159] Y. Chen, S. Barlowe, and J. Yang, "Click2annotate: Automated insight externalization with rich semantics," in *IEEE Symp. Visual Analytics Sci. Technol.*, Oct. 2010, pp. 155–162.

[160] L. Hong and E. H. Chi, "Annotate once, appear anywhere: Collective foraging for snippets of interest using paragraph fingerprinting," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2009, pp. 1791–1794.

[161] L. Hong, E. H. Chi, R. Budiu, P. Pirolli, and L. Nelson, "Spartag.us: A low cost tagging system for foraging of web content," in *Proc. Working Conf. Advanced Vis. Interfaces*, May 2008, pp. 65–72.

[162] N. H. Riche, B. Lee, and C. Plaisant, "Understanding interactive legends: A comparative evaluation with standard widgets," in *Proc. Comput. Graphics Forum*, vol. 29, no. 3, Jun. 2010, pp. 1193–1202.

[163] W. Willett, J. Heer, and M. Agrawala, "Scented widgets: Improving navigation cues with embedded visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1129–1136, Nov. 2007.

[164] R. V. Labaree. Research guides: Organizing your social sciences research paper: Writing field notes. [Online]. Available: http://libguides.usc.edu [Accessed: Dec. 12, 2017]

[165] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu *et al.*, "Galaxy zoo: Morphologies derived from visual inspection of galaxies from the sloan digital sky survey," *Monthly Notices Royal Astro. Soc.*, vol. 389, no. 3, pp. 1179–1189, Sep. 2008.

[166] Mikulski archive for space telescopes - stsci. [Online]. Available: archive.stsci.edu [Accessed: Feb. 2, 2019]

[167] A. S. Szalay, "The national virtual observatory," in *Proc. Astro. Data Anal. Software Syst. X*, vol. 238, Jan. 2001, p. 3.

[168] "The virtual astronomical observatory," http://www.virtualobservatory.org. [Online]. Available: http://www.virtualobservatory.org [Accessed: Nov. 7, 2018]

[169] G. Berriman, A. Laity, J. Good, J. Jacob, D. Katz, E. Deelman, G. Singh, M. Su, and T. Prince, "Montage: The architecture and scientific applications of a national virtual observatory service for computing astronomical image mosaics," in *Proc. Earth Sci. Technol. Conf.*, Jun. 2006.

[170] T. Boch, A. Oberto, P. Fernique, and F. Bonnarel, “Aladin: An open source all-sky browser,” in *Proc. Astronomical Data Anal. Software Syst. XX*, vol. 442, Jul. 2011, p. 683.

[171] M. Taylor, “Topcat: Working with data and working with users,” *arXiv preprint arXiv:1711.01885*, Nov. 2017.

[172] H. Dickinson, L. Fortson, C. Lintott, C. Scarlata, K. Willett, S. Bamford, M. Beck, C. Cardamone, M. Galloway, B. Simmons *et al.*, “Galaxy zoo: Morphological classification of galaxy images from the illustris simulation,” *Astrophys. J.*, vol. 853, no. 2, p. 194, Feb. 2018.

[173] A. Hassan and C. J. Fluke, “Scientific visualization in astronomy: Towards the petascale astronomy era,” *Publications Astro. Soc. Australia*, vol. 28, no. 2, pp. 150–170, Jun. 2011.

[174] D. R. Lipşa, R. S. Laramee, S. J. Cox, J. C. Roberts, R. Walker, M. A. Borkin, and H. Pfister, “Visualization for the physical sciences,” in *Proc. Comput. Graphics Forum*, vol. 31, no. 8, Dec. 2012, pp. 2317–2347.

[175] M. J. Turk, B. D. Smith, J. S. Oishi, S. Skory, S. W. Skillman, T. Abel, and M. L. Norman, “yt: A multi-code analysis toolkit for astrophysical simulation data,” *Astrophys. J. Suppl. Ser.*, vol. 192, no. 1, p. 9, Dec. 2010.

[176] R. Kaehler, O. Hahn, and T. Abel, “A novel approach to visualizing dark matter simulations,” *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2078–2087, Oct. 2012.

[177] K. Schatz, C. Müller, M. Krone, J. Schneider, G. Reina, and T. Ertl, “Interactive visual exploration of a trillion particles,” in *Proc. IEEE 6th Symp. Large Data Anal. Vis.*, Oct. 2016, pp. 56–64.

[178] K. R. Almryde and A. G. Forbes, “Halos in a dark sky: Interactively exploring the structure of dark matter halo merger trees,” in *Proc. IEEE Sci. Vis. Conf.*, Oct. 2015, pp. 73–77.

[179] M. Magnor, G. Kindlmann, C. Hansen, and N. Duric, “Reconstruction and visualization of planetary nebulae,” *IEEE Trans. Vis. Comput. Graphics*, vol. 11, no. 5, pp. 485–496, Jul. 2005.

[180] S. Wenger, M. Ament, S. Guthe, D. Lorenz, A. Tillmann, D. Weiskopf, and M. Magnor, “Visualization of astronomical nebulae via distributed multi-gpu compressed sensing tomography,” *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2188–2197, Oct. 2012.

[181] H. Li, C.-W. Fu, and A. Hanson, “Visualizing multiwavelength astrophysical data,” *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1555–1562, Oct. 2008.

[182] A. Bock, E. Axelsson, C. Emmart, M. Kuznetsova, C. Hansen, and A. Ynnerman, “Openspace: Changing the narrative of public dissemination in astronomical visualization from what to how,” *IEEE Comput. Graphics Appl.*, vol. 38, no. 3, pp. 44–57, May 2018.

[183] A. Sagrista, S. Jordan, T. Müller, and F. Sadlo, “Gaia sky: Navigating the gaia catalog,” *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 1070–1079, Aug. 2019.

[184] H. Li, C.-W. Fu, Y. Li, and A. Hanson, “Visualizing large-scale uncertainty in astrophysical data,” *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1640–1647, Nov. 2007.

[185] T. B. Luciani, B. Cherinka, D. Oliphant, S. Myers, W. M. Wood-Vasey, A. Labrinidis, and G. E. Marai, “Large-scale overlays and trends: Visually mining, panning and zooming the observable universe,” *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 7, pp. 1048–1061, Mar. 2014.

[186] P. Hanula, K. Piekutowski, C. Uribe, K. Almryde, A. Nishimoto, J. Aguilera, and G. E. Marai, “Cavern halos: Exploring spatial and nonspatial cosmological data in an immersive virtual environment,” in *Proc. IEEE Sci. Vis. Conf.*, Oct. 2015, pp. 87–99.

[187] S. Haroz, K.-L. Ma, and K. Heitmann, “Multiple uncertainties in time-variant cosmological particle data,” in *IEEE Pacific Vis. Symp.*, Mar. 2008, pp. 207–214.

[188] A. Preston, R. Ghods, J. Xie, F. Sauer, N. Leaf, K.-L. Ma, E. Rangel, E. Kovacs, K. Heitmann, and S. Habib, “An integrated visualization system for interactive analysis of large, heterogeneous cosmology data,” in *Proc. IEEE Pacific Vis. Symp.*, Apr. 2016, pp. 48–55.

[189] A. A. Goodman, “Linking visualization and understanding in astronomy,” in *Amer. Astro. Soc. Meeting Abstr. # 223*, vol. 223, Jan. 2014.

[190] C. Beaumont, T. Robitaille, A. Goodman, and M. Borkin, “Multidimensional data exploration with glue,” in *Proc. 12th Python Sci. Conf.*, Jan. 2013, p. 8.

[191] D. T. A. McGlynn, “Nasa fits.” [Online]. Available: fits.gsfc.nasa.gov [Accessed: Jul. 23, 2018]

[192] M. Turk, “yt.” [Online]. Available: yt-project.org [Accessed: May 25, 2018]

[193] D. Ceverino, A. Klypin, E. S. Klimek, S. Trujillo-Gomez, C. W. Churchill, J. Primack, and A. Dekel, “Radiative feedback and the low efficiency of galaxy formation in low-mass haloes at high redshift,” *Monthly Notices Royal Astro. Soc.*, vol. 442, no. 2, pp. 1545–1559, Aug. 2014.

[194] A. Zolotov, A. Dekel, N. Mandelker, D. Tweed, S. Inoue, C. DeGraf, D. Ceverino, J. R. Primack, G. Barro, and S. M. Faber, “Compaction and quenching of high-z galaxies in cosmological simulations: Blue and red nuggets,” *Monthly Notices Royal Astro. Soc.*, vol. 450, no. 3, pp. 2327–2353, Jul. 2015.

[195] M. Huertas-Company, J. Primack, A. Dekel, D. Koo, S. Lapiner, D. Ceverino, R. Simons, G. Snyder, M. Bernardi, Z. Chen *et al.*, “Deep learning identifies high-z galaxies in a central blue nugget phase in a characteristic mass range,” *Astrophys. J.*, vol. 858, no. 2, p. 114, May 2018.

[196] P. Jonsson, “Sunrise: Polychromatic dust radiative transfer in arbitrary geometries,” *Monthly Notices Royal Astro. Soc.*, vol. 372, no. 1, pp. 2–20, Oct. 2006.

[197] C. Y. Peng, L. C. Ho, C. D. Impey, and H.-W. Rix, “Detailed decomposition of galaxy images. ii. beyond axisymmetric models,” *Astro. J.*, vol. 139, no. 6, p. 2097, Apr. 2010.

[198] N. Elmqvist, P. Dragicevic, and J.-D. Fekete, “Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation,” *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 6, pp. 1539–1148, Oct. 2008.

[199] S. Liu, B. Wang, J. J. Thiagarajan, P.-T. Bremer, and V. Pascucci, "Visual exploration of high-dimensional data through subspace analysis and dynamic projections," in *Proc. Comput. Graph. Forum*, vol. 34, no. 3, Jun. 2015, pp. 271–280.

[200] H. Sanftmann and D. Weiskopf, "3d scatterplot navigation," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 11, pp. 1969–1978, Jan. 2012.

[201] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2376–2385, Oct. 2013.

[202] M. Gleicher, "Considerations for visualizing comparison," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 413–423, Aug. 2017.

[203] C. Henze, "Feature detection in linked derived spaces," in *Proc. IEEE Vis. Conf. '98 (Cat. No. 98CB36276)*, Oct. 1998, pp. 87–94.

[204] N. McCurdy and P. Moran, "Interactive exploration of solar magnetic field lines," in *(Presented at) 2017 Supercomputing Conference, Denver CO*, Nov. 2017.

[205] I. Hotz, "Visualizing line-like features – cleaning up the mess," in *(Presented at) The University of Utah. Salt Lake City, Utah*, Jan. 2019.

[206] J. J. Fang, S. Faber, D. C. Koo, A. Rodríguez-Puebla, Y. Guo, G. Barro, P. Behroozi, G. Brammer, Z. Chen, A. Dekel *et al.*, "Demographics of star-forming galaxies since z 2.5. i. the uvj diagram in candels," *Astrophys. J.*, vol. 858, no. 2, p. 100, May 2018.

[207] H. Zhang, J. R. Primack, S. Faber, D. C. Koo, A. Dekel, Z. Chen, D. Ceverino, Y.-Y. Chang, J. J. Fang, Y. Guo *et al.*, "The evolution of galaxy shapes in candels: From prolate to discy," *Monthly Notices Royal Astro. Soc.*, vol. 484, no. 4, pp. 5170–5191, Apr. 2019.

[208] C. Vertan, M. Ciuc, C. Fernandez-Maloigne, and V. Buzuloiu, "Browsing image databases by 2d image similarity scatter plots: Updates to the iris system," in *Proc. Int. Conf. Commun.*, Aug. 2002, pp. 397–402.

[209] N. A. Grogin, D. D. Kocevski, S. Faber, H. C. Ferguson, A. M. Koekemoer, A. G. Riess, V. Acquaviva, D. M. Alexander, O. Almaini, M. L. Ashby *et al.*, "CANDELS: The cosmic assembly near-infrared deep extragalactic legacy survey," *Astrophys. J. Suppl. Ser.*, vol. 197, no. 2, p. 35, 2011.

[210] J. Primack, A. Dekel, D. Koo, S. Lapiner, D. Ceverino, R. Simons, G. Snyder, M. Bernardi, Z. Chen, H. Domínguez-Sánchez *et al.*, "Deep learning identifies high-z galaxies in a central blue nugget phase in a characteristic mass range," *Astrophys. J.*, vol. 858, no. 2, p. 114, 2018.

[211] T. Munzner, *Visualization Analysis and Design*. Natick, MA, USA: AK Peters/CRC Press, 2014.

[212] M. Meyer and J. Dykes, "Criteria for rigor in visualization design study," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 87–97, Aug. 2019.

[213] H. A. Simon, *The Sciences of the Artificial*. Cambridge, MA, USA: MIT Press, 1996.

[214] G. Goldkuhl, "The it artefact: An ensemble of the social and the technical?–a rejoinder: An ensemble of the social and the technical?–a rejoinder," *Syst., Signs & Actions*, vol. 7, no. 1, pp. 90–99, Sep. 2013.

[215] A. Hevner and S. Chatterjee, "Design science research in information systems," in *Design Research in Information Systems*. Boston, MA, USA: Springer, 2010, pp. 9–22.

[216] J. Zimmerman, J. Forlizzi, and S. Evenson, "Research through design as a method for interaction design research in HCI," in *Proc. Conf. Human Factors Comput. Syst. ACM SIGCHI*, Apr. 2007, pp. 493–502.

[217] J. Pierce, "On the presentation and production of design research artifacts in hci," in *Proc. Conf. Designing Interactive Syst.*, Jun. 2014, pp. 735–744.

[218] L. David. Distributed cognition (dcog), in learning theories. [Online]. Available: https://www.learning-theories.com/distributed-cognition-dcog.html [Accessed: Feb. 1, 2020]

[219] E. Hutchins, "The social organization of distributed cognition." in *Perspectives on Socially Shared Cognition*. Washington, DC, USA: Amer. Psychological Assoc., 1991, pp. 283 – 307.

[220] J. Hollan, E. Hutchins, and D. Kirsh, "Distributed cognition: Toward a new foundation for human-computer interaction research," *ACM Trans. Comput.-Human Interaction*, vol. 7, no. 2, pp. 174–196, Jun. 2000.

[221] K. Höök and J. Löwgren, "Strong concepts: Intermediate-level knowledge in interaction design research," *ACM Trans. Comput.-Human Interaction*, vol. 19, no. 3, p. 23, Oct. 2012.

[222] B. Gaver and J. Bowers, "Annotated portfolios," *Interactions*, vol. 19, no. 4, pp. 40–49, Jul. 2012.

[223] J. Bowers, "The logic of annotated portfolios: Communicating the value of'research through design'," in *Proc. Designing Interactive Syst. Conf.*, Jun. 2012, pp. 68–77.

[224] M. M. Kennedy, "Generalizing from single case studies," *Eval. Quart.*, vol. 3, no. 4, pp. 661–678, Nov. 1979.

[225] W. A. Firestone, "Alternative arguments for generalizing from data as applied to qualitative research," *Educational Researcher*, vol. 22, no. 4, pp. 16–23, May 1993.

[226] A. K. Shenton, "Strategies for ensuring trustworthiness in qualitative research projects," *Educ. Info.*, vol. 22, no. 2, pp. 63–75, Jan. 2004.

[227] D. F. Polit and C. T. Beck, "Generalization in quantitative and qualitative research: Myths and strategies," *Int. J. Nursing Stud.*, vol. 47, no. 11, pp. 1451–1458, Nov. 2010.

[228] G. Ryle, *The Concept of Mind*. London, UK: Hutchinson & Co Ltd., 1949.

[229] C. Geertz, *The Interpretation of Cultures: Selected Essays*. New York, NY, USA: Basic books, Oct. 1973, vol. 5019.

[230] N. K. Denzin, *Interpretive Interactionism*. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2001, vol. 16.

[231] T. A. Schwandt and T. A. Schwandt, *Dictionary of Qualitative Inquiry*. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2001.

[232] I. Holloway, *Basic Concepts for Qualitative Research*. Oxford, UK: Blackwell Science, 1997.

[233] J. G. Ponterotto, "Brief note on the origins, evolution, and meaning of the qualitative research concept thick description," *Qualitative Rep.*, vol. 11, no. 3, pp. 538–549, Sep. 2006.

[234] S. Larsson, "A pluralist view of generalization in qualitative research," *Int. J Res. Method Educ.*, vol. 32, no. 1, pp. 25–38, Apr. 2009.

[235] R. M. Emerson, R. I. Fretz, and L. L. Shaw, *Writing Ethnographic Fieldnotes*. Chicago, IL, USA: Univ. of Chicago Press, 2011.

[236] S. M. S. da Cruz, M. L. M. Campos, and M. Mattoso, "Towards a taxonomy of provenance in scientific workflow management systems," in *IEEE Congr. Services-I*, Jul. 2009, pp. 259–266.

[237] Y. Zhao, M. Wilde, and I. Foster, "Applying the virtual data provenance model," in *Proc. Int. Provenance Annotation Workshop*, May 2006, pp. 148–161.

[238] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo, "Vistrails: Visualization meets data management," in *Proc. SIGMOD Int. Conf. Manage. Data*, Jun. 2006, pp. 745–747.

[239] J. Matejka, T. Grossman, and G. Fitzmaurice, "Patina: Dynamic heatmaps for visualizing application usage," in *Proc. Conf. Human Factors Comput. Syst.*, Apr. 2013, pp. 3227–3236.

[240] H. Stitz, S. Luger, M. Streit, and N. Gehlenborg, "Avocado: Visualization of workflow–derived data provenance for reproducible biomedical research," in *Proc. Comput. Graph. Forum*, vol. 35, no. 3, Jun. 2016, pp. 481–490.

[241] W. Pike, J. Bruce, B. Baddeley, D. Best, L. Franklin, R. May, D. Rice, R. Riensche, and K. Younkin, "The scalable reasoning system: Lightweight visualization for distributed analytics," *Inf. Vis.*, vol. 8, no. 1, pp. 71–84, Jul. 2009.

[242] D. A. Schön, *The Reflective Practitioner: How Professionals Think in Action*. Evanston, IL, USA: Routledge, 2017.

[243] B. Shneiderman, *The New ABCs of Research: Achieving Breakthrough Collaborations*. New York, NY, USA: Oxford University Press, 2016.

[244] M. Polanyi, *The Tacit Dimension*. Chicago, IL, USA: Univ. of Chicago Press, 2009.

[245] S. Thompson and N. Thompson, *The Critically Reflective Practitioner*. New York, NY, USA: Macmillan International Higher Education, 2018.

[246] P. Jarvis, "Reflective practice and nursing," *Nurse Educ. Today*, vol. 12, no. 3, pp. 174–181, Jun. 1992.

[247] C. Frauenberger, J. Good, G. Fitzpatrick, and O. S. Iversen, "In pursuit of rigour and accountability in participatory design," *Int. J Human-Comput. Stud.*, vol. 74, pp. 93–106, Feb. 2015.

[248] J. Saldaña, *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2015.