Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes

Eric D. Ragan, Alex Endert, Jibonananda Sanyal, and Jian Chen

Abstract— While the primary goal of visual analytics research is to improve the quality of insights and findings, a substantial amount of research in provenance has focused on the history of changes and advances throughout the analysis process. The term, *provenance*, has been used in a variety of ways to describe different types of records and histories related to visualization. The existing body of provenance research has grown to a point where the consolidation of design knowledge requires cross-referencing a variety of projects and studies spanning multiple domain areas. We present an organizational framework of the different *types* of provenance information and *purposes* for why they are desired in the field of visual analytics. Our organization is intended to serve as a framework to help researchers specify types of provenance and coordinate design knowledge across projects. We also discuss the relationships between these factors and the methods used to capture provenance information. In addition, our organization can be used to guide the selection of evaluation methodology and the comparison of study outcomes in provenance research.

Index Terms— Provenance, Analytic provenance, Visual analytics, Framework, Visualization, Conceptual model

1 Introduction

Data visualization and visual analytics combine the power of visualization with advanced data analytics to help people to better understand data and discover meaningful insights. While the goal of visualization research is ultimately to improve the quality of insights and findings, analytic processes are complicated activities involving technology, people, and real work environments. Practical applications encounter problems that extend beyond the integration of any system's analytic models, processing power, visualization designs, and interaction techniques. Visualization systems must also support human processes, which often involve non-standardized methodologies including extended or interrupted periods of analysis, resource sharing and coordination, collaborative work, presentation to different levels of management, and attempts at reproducible analyses [92, 52, 42].

For these reasons, a substantial amount of research in the areas of visualization, data science, and visual analytics has been dedicated to supporting *provenance*, which broadly includes consideration for the history of changes and advances throughout the analysis process (e.g., [34, 73, 37, 21]). It is clear that the research community agrees on the importance of supporting provenance, and many scholars have developed tools and systems that explicitly aim to help analysts record both computational workflows (e.g., [21, 5, 71]) and reasoning processes (e.g., [26, 37]). For example, *VisTrails* tracks steps of the computational workflow during scientific data analysis and visualization, and then provides graphical representations of the workflow through a combination of node diagrams and intermediary visual outputs [5, 14]. Groth and Streefkerk [39] presented another example with a system for recording and annotating stages of view manipulations during a 3D molecule-inspection task. As another example, Del Rio and da Silva [22] designed *Probe-It* to keep track of the data sets that contributed to the creation of map visualizations. Focusing on the provenance of insights, Gotz and Zhou described how the *HARVEST* system records the history of semantic actions during business and financial analysis activities [37]. These are just a few examples from a large number of visual analytics tools designed to support provenance across a wide range of domains and for different purposes.

As the body of research and existing tools has grown, the community's knowledge of the many factors and goals relevant for effective provenance support has also broadened. However, the variety of perspectives can make it challenging to assess the specific aspects and purposes of provenance that are targeted by any particular project. The term, *provenance*, has been used in a variety of ways to describe different types of origins and histories. For example, the scientific visualization community, especially the simulation and modeling communities, often interpret provenance as the history of computational workflow (e.g., [34]), while other interpretations focus on the history of insights and hypotheses (e.g., [70]). Although many researchers proactively provide clear definitions and explanations of their foci in the provenance research, this does not entirely resolve the challenge of consolidating the variety of interpretations and research outcomes across projects. Different perspectives and applications of concepts become problematic for interpreting and coordinating outcomes from different provenance projects, for communicating ideas within the visualization community, and for allowing new-comers to clearly understand the research space. In our work, we analyzed the different perspectives of provenance that are most relevant to areas of visualization and data analysis.

Our goal in this paper is to organize the different *types* of provenance information and *purposes* for why they are desired in information visualization, scientific visualization, and visual analytics. We present an *organizational framework* as a conceptual model that categorizes and describes the primary components of provenance types and purposes. Further, we discuss the relationships between these factors and considerations when capturing provenance information. Our organizational framework is intended to help researchers specify types of provenance and coordinate design knowledge across projects. In addition, our organization can be used to guide the selection of evaluation methodology and the comparison of study outcomes in provenance research.

2 Existing Perspectives of Provenance

Analytic provenance is a broad and complex concept within the areas of information visualization, data analysis, and data science. In visual data analysis, the concept often includes aspects of the cognitive and interactive processes of discovery and exploration, and also the computational sequences and states traversed to arrive at findings or insights. Prior surveys have presented definitions, categorizations,

- *Eric D. Ragan is with Texas A&M University. E-mail: ragane@acm.org.*
- *Alex Endert is with Georgia Tech. E-mail: endert@gatech.edu.*
- *Jibonananda Sanyal is with Oak Ridge National Laboratory. E-mail: sanyalj@ornl.gov.*
- *Jian Chen is with University of Maryland, Baltimore County. E-mail: jichen@umbc.edu.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

open questions, and potential areas of opportunities for analytic provenance for visual data analysis (e.g., [34, 97, 39, 37, 70]). Our goal in this section is to provide an overview of existing perspectives of provenance to explain the rationale for our organizational framework.

2.1 Provenance in Workflow Management

Approaching the topic from the perspective of data science and computational workflows, Zhao, Wilde, and Foster [97] distinguished between two components of provenance: *prospective* and *retrospective*. By this classification, *prospective provenance* includes the steps of a workflow procedure, while *retrospective provenance* includes technical information about the execution environment and resource consumption. Also focusing on workflows, Freire et al. [34] adopted the terminology of *prospective* and *retrospective* provenance in their survey of provenance for computational tasks. The authors added *causality* and *user-defined information*, with *causality* including on important information about process sequencing and coordination, and *user-defined information* including annotations or other documentation that better explains relevant user decisions associated with the workflow. Building off of this framework, the authors discuss three core components of provenance management systems: capture mechanisms, representational models, and the infrastructure for provenance storage and access. While the survey primarily focuses on practical methods for saving and retrieving information about computational workflows, the authors' discussion points to how these three components can generalize to other exploratory tasks and visualizations.

Serra da Cruz, Campos, and Mattoso [87] presented a taxonomy of characteristics of provenance from a technical perspective in workflow management systems. The highest level of the organization includes data *subject*, *capture*, *storage*, and *access* properties. The *subject* of provenance data describes information about when the data was collected, what granularity or level of detail was captured, and the combination of steps and input data that together make up the workflow. A system's *capture* properties describe the technical mechanisms that a system uses to collect provenance data, and it includes information about the relationship between provenance models, computational software, the operating system, and the stages of the workflow. *Storage* properties detail information about how the provenance data is saved, and *access* properties describe methods for retrieving provenance data.

Different features and thus inconsistencies in scientific visualization data analysis can arise depending on the choices that are made during the data analysis stages. The data analysis results can have significant impacts on the interpretation of scientific content often coming from simulation and modeling stages. Tracking and providing means for users to evaluate the quality of data and features is critical for reproducible analysis of any scientific data.

Scientific methods, data flows, annotations, evaluation protocols, automated decision making, and querying constitute important components of scientific workflows and have been well studied (e.g., [21, 34]). *VisTrails* was one of the earliest visualization tools that formally proposed a focus on the analysis pipeline to maintain an awareness of variability, perform parallel analyses, assess results for pipeline dependence, and maintain a detailed record of the analytical provenance of the secondary data generated from the raw datasets [5]. While reproducibility has increased substantially after substantial and thoughtful designs in *VisTrails*, provenance capture and review remain as complicated and difficult topics.

Discussing provenance organization in *VisTrails*, Scheidegger et al. [85] developed a three-layer model (evolution, workflow and execution) to organize changes in the workflow process where relationships among individual workflows and run time information about a specific outcome are shown. In this way, early outcomes can be explored or reproduced, and new exploratory tasks can be branched.

Many projects have tackled issues with provenance across a number of different application domains. For example, Braun et al. [10] discussed security concerns for the provenance of sensitive data such as medical and financial information. In another project, Santos et al. [19] explored the trustworthiness level and the reproducibility of the geographical information in a provenance visualization tool for the Global Earth Observation System of Systems (GEOSS). Anand et al. [2] presented an interactive provenance browser for visualizing and querying data dependency graphs for scientific workflows. The tool provided the user different views of the provenance as well as provided a query language to explore complex graphs.

2.2 Follow the Data

Deeply associated with the provenance of states in workflows (both scientific and decisional) is the provenance of data in the system. Algorithms act upon data and generate more data. Data participates in various transformations in its life cycle right from its generation to its deletion. Simmhan et al. [90] proposed distinguishing between *process provenance* (focusing on the workflow execution and the execution environments) and *data provenance* (focusing on the creation and transformation of data). The W3C group on provenance have proposed a data model (PROV-DM) that identifies six core components of provenance information as it relates to data: entities and activities; derivations of entities from entities; agents responsibility for entities and activities; a mechanism for recording the history of provenance information itself (provenance of provenance); properties to connect related or redundant entities; and logical structures to store and organize members [66]. Innumerable complications have arisen regarding data properties and considerations for provenance capture, storage, and access (e.g., [87, 43, 71, 21]).

Hensley et al. [43] explored unobtrusive mechanisms to facilitate provenance traces for the collection and use of sensor data. While their application area was specific, their design focus was to stay out of the typical analysis workflows of their users. They facilitated querying of the provenance through the interactive visualization of derivation trees. Much of their provenance capture was based on the Core Provenance Library (CPL), which is a general purpose provenance tracing library [61]. While most provenance systems require users to interact with their data using specially designed tools, the philosophy behind CPL is to abstract the provenance capture into general purpose affordances which allow client applications to determine their level of disclosure of provenance.

To facilitate the exploration of these provenance traces, Macko et al. [60] also developed a node-link visualization tool named *Provenance Map Orbiter*. Usability aspects of *Provenance Map Orbiter* were compared to a more refined tool named *InProv* by Borkin et al. [8] focusing on interaction mechanisms for high-level sensemaking. Del Rio and da Silva [22] have also studied usability but approach it from a perspective of data quality in map-making and argue for the inclusion of provenance in map generation. Also emphasizing visual design elements along with the flow of data, Maguire et al. [62] studied methods for determining glyph design for effectively representing workflow visualizations of biological data.

Accurate provenance traces are essential because decisional workflows require structured information. In addition, provenance information is meant to be used to facilitate the derivation of insight or actionable information; thus, it is often necessary to also capture domain specific knowledge. This is illustrated by Howe et al. [45] in their description of the development of a provenance-aware workflow and 3D visualization system for data analysis tasks for oceanic applications.

Different approaches have been employed to deal with the data deluge common in provenance of workflows. Biton et al. [6] approached the overwhelming quantity of bioinformatics workflow data by controlling the visibility of sub-level workflow views, thus limiting the provenance information most relevant for the user. Also related to the quality of data history, Chen et al. [17] studied provenance visualization to study error propagation and compare among network traces. Focusing on representing and understanding the provenance data, Chen and Plale [16] explored layout algorithm, visual style, graph abstraction techniques, graph matching algorithm, and temporal representation technique to deal with the high complexity of large scale scientific data provenance.

2.3 Graphical History

More specific to visual analytics, many visualization provenance tools also record the intermittent visual outputs during an analysis (e.g., [27, 95, 57, 39, 22, 72]), providing many of the benefits of the well known *VisTrails* tool [5]. Perhaps focusing more on exploratory analysis, Heer et al. [41] use the term *graphical histories* to refer to history of visualization states, as presented with a graphical interface that allows saving and revisiting visualizations in data analysis with Tableau. Jankun-Kelly et al. [49] described the p-set model of visualization exploration to describe the history and derivation of the exploration of the visualization process. Javed and Elmqvist [50] also present a tool, *ExPlates*, that presents a spatial canvas of visualization states that serves as graphical history while allowing visual comparison of multiple views. Work with visual history by Ragan, Goodall, and Tung [77] and by Andrews, Endert, and North [3] studied how the spatial organization of visual artifacts in an analysis workspace affects sensemaking when screen space provides a persistent record of viewed information.

These are just a few of many examples that demonstrate the perspective of provenance of visualization state. While undoubtedly important and highly valued, the concept of supporting provenance of visualization states or visual history is relatively simple: save visualizations to help people remember and revisit earlier states in analysis.

2.4 Interaction History

In addition to recording the history of workflow, data, and visual representations, many researchers have emphasized the value of following the history of user interactions. Gotz and Zhou [37] used the term *insight provenance* to refer to the history of actions and rationale during an analysis process. By their explanation, insight provenance accounts for visual actions, the exploration of information, and the analytic insight gleaned from the analysis. Gotz and Zhou characterize different levels of tasks and actions as important components to consider for insight provenance, with levels covering high-level analysis objectives, more specific sub-tasks, and even more specific descriptions of interface interactions. An important distinction for actions for within this insight provenance framework is that the actions include *intentions* for taking those actions. The authors identified four types of intent: changing data, changing visual state, changing notes, and changing the action history. The researchers went further with a taxonomy of user actions with categories to distinguish among categories for *exploration actions* (which includes data and visual explorations), *insight actions* (those that help record insights, findings, or important information), and *meta actions* (those relating to command history, such as undo/redo).

A great deal of other work has focused on recording and visualizing the history of interactions (e.g., [54, 24, 65]). Cowley et al. [20] developed a system called *Glass Box* that records all low-level user interactions and events on the operating system of an analyst. They describe how the capture and storage of the interactions and events were successful, but the analysis and understanding of these logs remained an open challenge. Dou et al. [26] found that the manual human analysis of user interactions of another person's analysis can lead to understanding and recovering of specific aspects of the reasoning process. More recently, Brown et al. [12] showed how user interactions of a visual search task can be analyzed to find metrics about a user's performance, such as time to completion and success rate. Similarly, Fink et al. [33] found that *save* and *save as* marked important milestones for cyber security analysts during their investigations. From these works, we learn that user interaction does encode some aspects of the analytic process and reasoning, and thus makes for a valuable source of analytic provenance data.

Examples exist on how to integrate such interaction provenance into the interface and visualization during analysis (as compared to analyzing it post-hoc). For example, Matejka et al. [65] have shown how frequent use of menu items can be shown by the graphical depiction of the button or menu item appearing worn, or otherwise illuminated using a heatmap. Their work illustrates how usage counts for functionality can be represented to help users understand usage history, and over time adapt their interface to enhance productivity by creating more visually salient menu items. Similarly, work on model steering has shown how user interactions can be systematically analyzed to steer the computation for visual analytic tools [12, 29, 79, 30].

2.5 Sensemaking and Insight

Many researchers also discuss the importance of capturing user thoughts, analytical reasoning, and insights during analysis. Defining insight for visual data exploration is complex, and it is likely that a single agree-upon definition does not exist. For example, Chang et al. [15] present views and definitions about the term insight from different communities of science. The cognitive sciences have used the term to indicate a process by which a problem solver suddenly moves from a state of not knowing how to solve a problem to a state of knowing how to solve it. [64]. In the visualization community, insight often describes intermediate or final outcomes or findings that result from using the visualization [32]. The importance of insight to information visualization is also evidenced by techniques such as insight-based evaluation that can be used to evaluate the effectiveness of visualizations [82]. More recently, Stasko points out that in addition to the ability for insight to represent findings or knowledge, it may also include a notion of the additional questions about the data that were previously unknown to the analyst [91]. Thus, the *insights* area an important part of characterizing and understanding analytic provenance.

Xu et al. [96] discussed challenges for provenance of sensemaking, including issues such as uncertainty, semantic hierarchies, manual and automatic capture mechanisms, and approaches for visualization and presentation. North et al. [70] structured their discussion of analytic provenance through the lenses of perceiving, capturing, encoding, recovering, and reusing. They emphasize how this sequence is critical to creating holistic visual analytic applications that incorporate analytic provenance functionality. An important commonality is the need to capture aspects of provenance as the basis upon which other functionality can be performed (e.g., modeling, reuse, query).

User interaction data is one common form of data collected about analytic provenance, but others have included annotations, screenshots, and data transformations in related discussions. For example, Groth and Streefkerk [39] present a model for augmenting the systematic recording of series of visualization screenshots and user interactions with user-generated annotations. As a result, the automatically captured data is augmented with higher-level, semantic information from the user about his or her process. This aids in the subsequent uses of analytic provenance, such as recall.

Additionally, Heer and Shneiderman [42] discuss how the recording and capturing of analytic provenance information can foster benefits to data analysts including sharing the processes with others and guiding users through tasks or stories. They also emphasize how there are two forms of data to capture: designed systematic capture (e.g., user interactions, screenshots of visualization) and explicit user annotations.

These prior projects and categorizations of analytic provenance for visual data exploration ground the framework presented in this paper. We contribute to the knowledge in this community by organizing the multiple perspectives of provenance within the framework and discussing how it can be potentially used for different tasks and purposes.

3 Method

The presented framework is based on both formal and informal reviews of literature related to provenance in the fields of visualization and data science. The formal review covered 50 papers. To begin, we reviewed an initial set of 25 papers and used selective coding to take notes for the following:

- The terminology, definitions, and descriptions used by the authors to refer to elements that are relevant to provenance.

- A summary of the focal elements of the work as interpreted from the authors' terminology along with the full description of the work.
- The purposes and goals of provenance discussed or supported in the work.
- The disciplines or domain areas for the tools or systems presented in the paper.
- Other notes relevant to provenance, categorization, relevance, and focus.

We then used the notes from the selective coding of the initial set of papers to develop a preliminary set of categories.

Next, we added an additional 25 papers for selective coding, and we performed axial coding on the core set of 50 papers to organize the types and purposes of each paper using the initial categories. This was an incremental process of review and category revision. When we encountered concepts that did not fit the existing set of categories, we discussed the organization and revised the categories according to our best judgments. Following each category revision, an additional pass of axial coding was made over the selective codes.

In addition, while conducting our literature review and identifying relevant papers to include in the core paper set, we informally reviewed a greater corpus of papers. We used this broader review to help achieve appropriate breadth while determining the scope of our analysis (for example, because we focused on visualization, we limit coverage of provenance in database research). After establishing the version of the organizational framework presented in this paper, we continued reviewing additional papers to check our categories. Because these reviews were conducted informally, records of the process were not recorded.

Types of Provenance Information	
Data	The history of changes and movement of data, which can include subsetting, data merging, formatting, transformations, or execution of a simulation to ingest or generate new data
Visualization	The history of graphical views and visualization states
Interaction	The history of user actions and commands with a system
Insight	The history of cognitive outcomes and information derived from the analysis process, including analytic findings and hypotheses
Rationale	The history of reasoning and intentions behind decisions, hypotheses, and interactions

Purposes for Provenance	
Recall	Maintaining or recovering memory and awareness of the current and previous states of the analysis
Replication	Reproducing the steps or workflow of a previous analysis
Action recovery	Maintaining the action history that allows undo/redo operations and branching actions during analysis
Collaborative communication	Communicating and sharing data, information, and ideas with others who are conducting the same analysis
Presentation	Communicating the insights or progression of the analysis with those who are not directly involved with the analysis themselves, such as general public, upper levels of management, or analysts focusing on other areas
Meta-analysis	Reviewing the analytic processes themselves in order to understand and improve aspects of the analysis (such as process efficiency, training efficiency, or analytic strategies)

Table 1. Summary of the organizational framework of provenance types and purposes.

4 Organizational Framework

It is clear that different researchers and projects consider provenance from different perspectives in the area of visual analytics. In this section, we organize the perspectives for types of provenance information and the purposes for using that information. An overview of the framework is shown in Table 1.

4.1 Types of Provenance Information

Based on our review of the literature, visual analytics research tends to focus on five types of provenance information types: *data*, *visualization*, *interaction*, *insight*, and *rationale*. These categories are not perfectly disjoint classifications, as analytic workflows are inherently complex, and the type of provenance are interrelated. In fact, it would be difficult to capture a single type without also collecting information about others; however, many projects do emphasize capture of certain forms of provenance information.

4.1.1 Provenance of Data

The *provenance of data* focuses on the history of changes and movement of data. Data provenance is often heavily emphasized in computational simulations and scientific visualization, in which significant data processing is conducted (e.g., [21, 43, 97]). The history of data changes can include subsetting, data merging, formatting, transformations, or execution of a simulation to ingest or generate new data. From the visualization perspective, workflow and data are closely related and descriptive of the technical facets of provenance in visual analytics. Therefore, in our organization, we also include information about workflow execution, services, and computational environments under the data type of provenance. Thus, as a high-level type of provenance type, provenance of data includes both the *prospective* and *retrospective* provenance types described by Zhao, Wilde, and Foster [97] as well as both the *process provenance* and *data provenance* types described by Simmhan et al. [90].

The capture of provenance related to data is complex and its relation to provenance is probably most direct. While actions on data in workflows are prospective and hooks into code that allow the capture are usually directly encoded, other steps in the workflow may be much more difficult to capture. Freire [34] described how capture can be be based on different levels, such as via workflow, process, or operating system levels. In addition, the provenance of data often involves challenges of versioning and forking, updates to data sets, movement, duplication, and associated levels of uncertainty. Quality control and assurance are well structured; therefore, their application on data can be easily facilitated, and both prospective and retrospective mechanisms are effective for capture.

On the other hand, data often contains provenance information as meta information when there is no centralized management of this information. Provenance systems with decoupled provenance information require managed handlers for the data, such as in *SPADE* [80]. There have also been arguments for including provenance capabilities built into the operating system itself (e.g., [44, 75, 23]).

4.1.2 Provenance of Visualization

Rather than focusing on computation and the data itself, visualization provenance is concerned with the history of graphical views and visualization states. In practice, the history of visualization states is tightly coupled with data transformation and the interactions used to produce the visualization; however, it is important to realize that it is possible to record one of these provenance types without the others. Many visual analytics tools emphasize support for visualization provenance. Examples include *Chimera* [54], *VisTrails* [5], *GraphTrail* [27], the *Graphical History Interface* by Heer et al. [41], the *TimeTravel* interface by Derthick and Roth [24], and the 3D timeline view by Dobo, Mitra, and Steed [25].

Capturing the provenance of visualization is relatively straight forward: the system needs to save either an image of the visualization or the state and settings needed to recreate it later. Another question is how much source data or state information is necessary to save along with the visualization information. More data may be beneficial for understanding the context of creation or branching modified visualizations, but such benefits come with storage costs.

4.1.3 Provenance of Interactions

The *interaction provenance* type focuses on the history of user actions and commands with a system. Rather than individual perspectives of system events or human cognitive processes, interactions include explicit and observable user interactions between the two. By this interpretation, we are separating the action and rationale components of insight provenance as described by Gotz and Zhao [37]; our *interaction provenance* category includes actions, while our *rationale provenance* category includes intentions. While closely related, the distinction is important for the consideration of capture mechanisms because while interactions can be clearly observed and automatically captured, intentions and rationale are internal cognitive processes.

However, our perspective of interaction provenance does cover the action space described by Gotz and Zhou [37]. *Data exploration interactions* include operational actions that aid in viewing the data, with examples including button pushes, view manipulations, and query executions. Exploration actions can also be considered to include the common types of analytic tasks as described by other taxonomies, such as the analytic tasks described by Amar, Eagan, and Stasko [1], the user objectives described by Wehrend and Lewis [93], or the actions and operator primitives described by Roth [78]. Task-types in these taxonomies include analytic actions such as search, comparison, clustering, ordering, and filtering. *Annotation interactions* include actions and inputs that provide supplemental user information about the nature of data manipulations, intentions, rationale, or insights during the analysis (note that though the observable actions for supplying this information are included under interaction provenance, the types of information being supplied may be categorized differently). In addition, *command history actions*, such as undo, redo, bookmarking, and step re-tracing allow users to revise, reassess, or revisit the history of previously taken actions.

While interactions performed within this category may effect a change in data or help capture user thoughts, we reiterate that the focus is on execution of actions, and the necessary interactions with the technology are the primary emphasis of the *interaction provenance* perspective. Because the interactions are technologically based, the history of actions can be captured automatically through system logs, as it is not difficult to record button presses, mouse cursor movements, or even physical movements [4, 65, 13]. However, practical decisions on how to capture sufficient interaction data at the appropriate resolution can be challenging. Insufficient level of detail could be a major problem, but a record of absolutely every user interaction could be difficult to process and interpret. It can also be difficult to know in advance which types of interactions might be most valueable later on.

4.1.4 Provenance of Insights

Insight provenance includes cognitive outcomes and information derived from the analysis process. This category includes the history of hypotheses, insights, and other forms of analytic findings due to data exploration and inference.

Unfortunately, unlike the history of interactions or data computations, insights are not directly observable. For evaluation purposes, benchmark tasks have proven useful for identifying and assessing analytic findings [74], but insight recognition can be challenging in more realistic settings. North [69] commented on the challenges of identifying and capturing insight for the purpose of evaluation. He argues that open-ended and qualitative methods are necessary for recognizing instances of insight. Often, visual analytics systems that support insight provenance include capture methods that require users to provide additional information. This is often done through annotations to record important findings (e.g., [72, 94]), recording and analyzing verbal expressions (e.g., via think-aloud protocol), [77, 58] or physiometric responses (e.g., eye tracking, electroencephalogram) [48, 59, 9].

4.1.5 Provenance of Rationale

For a deeper understanding of insights and interactions, it is necessary to understand the history of user intentions and reasoning behind them. *Provenance of rationale* strives to capture the reasoning behind decisions, hypotheses, and interactions. This type of provenance information goes beyond specific intentions for individual changes made to data, visualization, or annotations, and it also includes higher-level objectives that motivate analysis (as described by Gotz and Zhou [37] and Roth [78]). Ideally, a complete record of reasoning will elucidate a user's analytic strategy.

Like capturing insight provenance, recording rationale is generally not directly possible without collecting additional user information. Furthermore, while insights usually occur relatively infrequently, thought and rationale are continually ongoing throughout an analysis. As a result, capturing a complete record of rationale is difficult (or perhaps impossible). Constantly providing updates on thoughts and rationale can invoke an overhead cost to the analysis or detract from the analysis activity [7, 31]. On the other hand, a study by Dou et al. [26] found evidence that a significant amount of reasoning information can be inferred by analyzing system logs of events and interactions. Anecdotally from our own experiences, we know it is often easiest to infer reasoning from system logs when events follow a normal sequence of progression, and additional annotation is most import for explaining unexpected occurrences or divergence from the norm. Regardless, it is important for provenance tools to support capturing rationale, and many visual analysis tools do (e.g., [58, 72, 94]).

4.2 Purposes for Provenance

Related to the types of provenance information is how that information will be used. Despite the consensus in the visualization community that provenance information is beneficial, different projects emphasize benefits for different purposes for its use. At the highest level, it is important to clarify whether types of information are intended to be used to support *ongoing analysis purposes* or for *post-hoc purposes* after an analysis is completed. At another level of specificity, we can organize purposes under *recall*, *replication*, *action recovery*, *collaborative communication*, *presentation*, and *meta-analysis*.

4.2.1 Recall

In one form or another, the purpose of supporting *recall* of an analysis or workflow can be seen the majority of provenance efforts in visual analytics (e.g., [77, 42, 50, 58, 5]). Recall of the analytic process is perhaps the most generic purpose for provenance. Recall can be thought of as a component of other purposes, such as replication, presentation, or collaborative communications. But maintaining process memory and recall is also important on its own during analysis. Recall enables awareness and understanding of what analytic tasks have previously been completed, what findings have been established, and what tasks remain to be completed in the future. Thus, recall is important for analytic clarity and efficiency. The benefits of process recall become even more important over extended or intermittent periods of analysis, in which it can be harder to remember previous analytic states [77].

4.2.2 Replication

Another common purpose for using provenance information is to reproduce a previous analysis (e.g., [21, 51, 71, 97]). *Replication* can be thought of as the application of recall after the analysis in order to repeat or verify results. Other times, an analytic reproduction can provide a basis for branching investigations or revised analyses with modification of parameters. Replication is important for validation and continued analysis after a previous analysis, and it can also enable more thorough exploration of possibilities during an analysis.

4.2.3 Action Recovery

Support for *action recovery* involves maintaining an action history that allows undo/redo operations and branching actions during an analysis. While a simple purpose, action recovery is critical for enabling exploratory analyses, providing resilience to human errors, and enabling transitioning to a previous state of the analysis. The ability to recover from errors is a fundamental component of usability [68] and is commonly supported in visual data exploration applications via numerous methods, such as undo/redo, bookmarking, or timeline views [24, 25, 11]. As the purpose is to recover from actions and leverage command history, it is no surprise that *action recovery* is most closely connected to the *interaction* type of provenance, but this is not necessarily the only relevant information type.

4.2.4 Collaborative Communication

Collaborative communication involves communication with others who are conducting the same analysis. Collaboration may be conducted synchronously or asynchronously and in collocated or distributed settings. Purposes for collaboration are similar to those described for recall but with the added complication of helping other people to understand the state of the analysis. This is important for establishing common ground among analysts. The importance of provenance information in collaborative settings will depend on the degree of coupling between individual analytic activities and the complexity of the analysis. Many collaborative analytic tools support the sharing and communication of ideas and information, such as through shared annotations, brushing and linking, or shared activity indicators (e.g., [63, 18, 46, 67, 35, 28]).

4.2.5 Presentation

To distinguish between collaborative communication, *presentation* often involves communication with those who are not directly involved with the analysis themselves. Examples include reporting to the general public or to upper levels of management, audit reports, or teaching. By this view, presentation is generally not performed in the midst of an ongoing presentation. Presentations are important for communicating how an analysis was conducted, how the findings were determined, or how the data justifies a conclusion. Such purposes are closely related to the areas of storytelling or narrative visualizations (e.g., [84, 53, 86, 36]). Provenance information may be important for communicating an accurate and coherent story.

4.2.6 Meta-Analysis

Other purposes for provenance involve reviewing the analytic processes themselves. *Meta-analyses* of processes make it possible to review and evaluate process efficiency and understand analytic strategies. This information can be used to improve systems or approaches to analysis. For example, understanding computational bottlenecks can help inform decisions for prioritizing upgrades or focal points for workflow optimization. In exploratory analyses, the identification of strategies or biases can be used to help analysts improve their methods or to train new analysts. Meta-analysis purposes for using provenance information are most likely useful post-analysis; however, analytic methods could make it possible to assess analysis patterns and help optimize performance or suggest changes in real time during analysis.

5 Discussion

Organizing provenance information by types and purposes provides benefits for informing decisions for capture mechanisms, necessary levels of granularity, and the appropriate means of evaluation. In addition, the organizational framework provides a convenient scaffold for comparison and cross-analysis of tools and studies.

5.1 Capture, Granularity, and Uncertainty

Understanding the type of provenance information of interest and how it will be used can help inform the decision for how to capture the provenance information. Thus, the type of information strongly relates to the appropriate means of collection. The provenance of data, visualizations, and interactions can often be collected automatically by saving system logs for events, system inputs, and system outputs; however, such methods do not directly capture the provenance of insights or rationale. On the other hand, insights and rationale can sometimes be inferred from the more system-oriented information. We posit that when describing an application or system that supports provenance, it is important to clarify what provenance information is being recorded explicitly, what information can be inferred, and what types may not be captured at all.

In practical situations, it is not easy to completely capture all types of provenance information. Provenance information can be captured with varying levels of detail, and tracking changes and processes can become more challenging in scenarios with heterogeneous data types and sources. Consideration for the intended purpose of the saved information can inform decisions for prioritizing capture mechanisms. Further, understanding purposes for use can help designers to determine the level of *granularity* of provenance capture. Provenance granularity describes the level of detail of the captured information (e.g., [87, 56, 54, 89]). Reduced granularity can correspond to greater uncertainty about underlying aggregations of data, processes, or actions when interpreting the information. While it may be desirable to capture as much detail as possible, this can be difficult and invoke significant requirements for data storage. Further, additional costs may be met for accessing or interpreting the provenance information. Depending on how the information will be used, it may not be necessary to capture as much detail as possible. As a simple example, one way to capture interaction provenance with a standard computer application would be to record every keystroke, every mouse click, and the position of the mouse cursor at every millisecond. The benefits of such detailed granularity over recording higher-level interactions, such as query executions or interface button clicks, is questionable for most practical purposes. In other situations, such as when considering the history of data or visualization transformations, it can be more complicated to determine how the level of granularity will influence effective use.

Uncertainty of information and the uncertainty of the provenance itself may creep into the system when appropriate affordances to explore the provenance are not available or are unsupported by the provenance system. Design requirements for the provenance capture system and the provenance exploration system should be considered together, as they are complementary; however, implementation requirements can be addressed separately. This facilitates building extensible analytics modules over the appropriate level of granularity for provenance capture. It is important that users understand the limitations and associated uncertainty of captured provenance information. Careful design can help mitigate systemic uncertainties in the interpretation of provenance. This is illustrated in Hensley [43], where interactive graph views facilitate the exploration of provenance at different levels of granularity. Chen et al. [17] explored yet another aspect of provenance uncertainty: comparison of provenance graphs that facilitate the exploration of how repetitions of actions differ when they are re-enacted. Visualization techniques for communicating uncertainty can borrow from the body of uncertainty visualization literature (e.g., [38, 81, 40]), but additional research is needed to understand the effectiveness of uncertainty representations within the context of history and workflow visualizations.

5.2 Interpreting Types and Purposes in Existing Work

The organizational framework presented in this paper provides a method for analyzing and comparing visual analytics applications and projects. Table 2 demonstrates how the framework can be used to describe and compare different work that supports information history and provenance. Each row of the table corresponds to a publication, and the columns correspond to areas of emphasis in each body of work. Note that the amount of types or purposes emphasized is not necessarily related to quality of the system or research outcome, as it is logical for most applications to focus on a specific purpose suitable in its application domain. That is, in some situations, it may not be necessary to capture all types of provenance information types, and that information may only need to be used in a limited number of ways. We also note that the distinction of types of purposes for any given project are often open to some interpretation. In addition, we mention that the levels of emphasis are based on the described capture methods and uses focused on in the papers, rather than all purposes for how the system or provenance information could potentially be used.

Table 2 was created from a subset of the coded papers (as described in section 3) to summarize how different papers contributed to the categories of the framework. The levels of emphasis and the

Citation (Abbreviated)	Discipline/ application	Type Emphasis: Data	Vis	Interaction	Insight	Rationale	Purpose Emphasis: Recall	Replicate	Action Recovery	Collab. Comm.	Present	Meta-Analysis
Bavoil 2005 [5]	scientific workflows	dark	dark	light	light		dark	dark	dark	dark		
Del Rio 2007 [22]	map creation	dark	dark				dark	dark	light			light
Derthick 2001 [24]	various		dark	dark			dark		dark			
Doboš 2014 [25]	3D modeling	dark	dark	dark			dark	dark	dark			
Dou 2009 [26]	financial analysis			dark	light		dark			light	light	
Dunne 2012 [27]	network graph		dark	dark			dark		dark	dark	light	light
Ellkvist 2008 [28]	scientific workflows	dark	light	light			dark	dark	light	dark		
Gotz 2008 [37]	business; financial			dark	dark	light	dark		dark		light	
Groth 2006 [39]	3D molecule vis		dark	dark	dark	dark	dark		dark			
Heer 2008 [41]	various		dark	dark	light	light	dark		dark	dark	dark	
Heer 2012 [42]	various		dark	dark	dark	dark	dark	dark	dark	dark	dark	light
Hensley 2014 [43]	sensor data workflow	dark					light	dark				dark
Javed 2013 [50]	various	dark	dark	light	dark	dark	dark	light				light
Kadivar 2009 [51]	various		dark	light	light	light	dark	dark				light
Kurlander 1988 [54]	graphics editor		dark	dark			light		dark			
Lipford 2010 [58]	financial analysis			dark	light		dark			light	light	
Maguire 2012 [62]	scientific workflow	dark					dark	dark				light
Mahyar 2014 [63]	intelligence analysis				dark	dark	dark			dark		
North 2011 [69]	various			dark	dark	light	light	dark				
Parker 2005 [71]	scientific workflows	dark										dark
Shrinivasan 2008 [89]	various	dark	dark	dark	dark	light	dark	dark	dark		light	
Simmhan 2006 [90]	scientific workflows	dark		light	light		dark	dark		light		

Table 2. Overview of perspectives of provenance information types and purposes as emphasized in a sample of visualization projects. Provenance information types are shown in the blue columns, and purposes are shown in the purple columns. Darker colored cells indicate heavier emphasis, and white cells indicate low emphasis.

corresponding color codings were subjectively derived from our notes at the conclusion of the iterative review process. Obviously, the table is far from a comprehensive list of relevant papers; its purpose is to demonstrate the variety of focal areas of provenance research in the realm of visualization. With this high-level view, it is possible to easily distinguish focal areas and compare goals of different systems and efforts. For example, the table shows that heavy emphasis on recall is common among the collection of papers. This should not be surprising since recall is the most fundamental and inclusive type of purpose in our organization.

The construction of Table 2 also demonstrates how characterizing specific provenances types and purposes can be difficult in practical scenarios, as the complexity of real situations and the interconnected nature of various types and purposes are unavoidable. Despite this limitation, the framework can still be used as a starting point when describing goals and methods when supporting provenance.

This may be useful when comparing work or for identifying open research opportunities. For example, our review of the literature revealed few projects that heavily focus on using provenance to support presentation or meta-analysis. While several projects do serve as examples (see Table 2), work is particularly limited for supporting such purposes in projects focusing on provenance of rationale. This suggests that these research areas could benefit from further development. Capture and review of analytic rationale could help to improve analytic strategies and methods, and tools that leverage the benefits of narrative and presentation could assist the presentation of rationale. The advancement of visualization designs and the execution of new evaluation efforts could be highly beneficial in these areas.

On the other hand, our review and the sample in Table 2 indicate that substantial work exists involving the provenance of data, visualizations, and interactions. Similarly, research support has been relatively strong for the purposes of recall, replication, and action recovery. Of course, this does not mean that these areas are ”solved problems”. For example, in our review, we found a limited amount of work providing thorough and convincing evaluations of provenance tools. We therefore recommend additional research progress in evaluating the effectiveness of provenance systems for different purposes.

5.3 Guiding Evaluation of Provenance Support

Taking the ability to support comparison a step further, the framework can also be used to compare results of different evaluations of provenance tools. Evaluation is important for understanding the effectiveness of provenance tools and how to improve them. In order to take advantage of existing design knowledge and advance effective new designs, it is not sufficient for evaluations to simply conclude that a system does or does not support provenance. Evaluations rely on metrics that are based on goals, tasks, and purposes. By clarifying the purpose for provenance use that is evaluated, it is possible to derive a more specific body of design knowledge. Thinking in terms of provenance types and purposes can help guide the design of evaluations to appropriately assess the effectiveness and appropriateness of different tools and designs. Furthermore, our organizational framework can help to compare the findings of different evaluations.

Researchers employ various approaches to evaluating visualization tools [55, 74]. One way of determining whether a tool is effective is to ask real professionals or experts to use the tool for an extended period of time to do their work. Observations and interviews allow researchers to understand the ways that the tool is most helpful and to identify weaknesses. In addition, logs of tool usage can capture performance information, identify problems that users might have with completing their tasks, or reveal patterns in how different tool funcationalities are used (e.g., [26]). Such evaluation methods can also be used in controlled experiments, in which study participants (who are not necessarily experts) use the tools for short periods of time to complete focused tasks. Controlled experiments allow targetted comparisons of how different tool features and functions affect behaviors and performance outcomes [76]. In addition to user studies, heuristic evaluations and cognitive walkthroughs can be used to identify problems or assess suitability of a system for supporting particular tasks and objectives. For all of these evaluation options, it is important to be able to clearly describe the objectives and the design factors being evaluated.

It will become increasingly important to maintain a set of categories of provenance types and purposes as evaluation research in this area advances. Ragan and Goodall [76] discussed evaluation methodology for eliciting and measuring the quality of recall and communication of the provenance of interactions, rationale, and insights through controlled studies. This methodology covers approaches for quantifying performance outcomes specialized to match different purposes for provenance (e.g., replication, communication, and recall). Lam et al. [55] categorized various approaches to conducting empirical studies of visualization. One such category of studies, *visual data analysis and reasoning* (VDAR), is highly relevant for provenance evaluation because its goal is to "assess a visualization tools ability to support visual analysis and reasoning about data" [55]. The authors empahsize that VDAR studies can not only evaluate performance metrics, but they should also address how well the tools support the analytic process as a whole. Our framework can be used to help clarify how a study supports the overall process by encouraging specificity when describing provenance-centric purposes. At present, the number of rigorous evaluations of provenance tools is limited. A recent review of visualization evaluations by Isenberg et al. [47] found that few evaluation papers (2.9% of the review sample) fell into the VDAR category, which would cover data analysis processes and provenance management. Furthmore, the sample included no evaluations that assessed commmunication through visualization, which we argue is an important purpose of provenance systems. These results complement our review's finding that a relatively limited amount of research covers the provenance of rationale and the presentation of provenance information.

Of the existing evaluations of provenance systems and visualizations, different studies focus on different aspects. As one instance, Ragan et al. [77] evaluated how the level of visual detail of the analysis workspace affected recall of analysis. In a study of provenance exploration tools for collaborative analysis, Sarvghad and Tory [83] compared different tools that helped participants to understand other analysts' histories. As another example, Lipford et al. [58] performed a study to evaluate how well *WireVis*, a visualization system for analysis of financial data, helped users recall insights, interactions, and rationale.

Other comparative studies have focused on different uses for provenance. For instance, Groth and Streefkerk [39] studied how different action recovery techniques affected task performance in a spatial inspection task, and Del Rio and da Silva [22] evaluated a provenance visualization tool showing the provenance of maps with a visual history of workflow and data sources for different maps. In addition, researchers haved studied how analysts use different types of provenance features through field studies with analysts (e.g., [27]) or by analyzing system logs (e.g., [41]). With numerous studies and findings, having a framework to organize provenance types and purposes will facilitate comparisons of the purposes and outcomes of such studies in meaningful ways.

Such an organization will become increasingly important as design knowledge continues to advance for provenance support systems. Thus, moving forward, we recommend that researchers consider *purposes for provenance use* as additional types of *analytic tasks*, similar to how researchers refer to taxonomies of analytic tasks or objectives. Many variations of task taxonomies exist to help organize different types of analysis goals in visual analytics (e.g., [1, 93, 88, 78]). Despite some variation in specific terminology and scope, these organizations include tasks such as search, identification, comparison, clustering, ordering, filtering, pattern detection, and anomaly detection. We argue that provenance uses (e.g., recall, replication, action recovery, collaborative communication, presentation, process meta-analysis) should also be considered as fundamental tasks for visualization use.

6 Conclusion

Provenance is an important topic in visualization research, but its meaning and focus can be interpreted in different ways. We present an organizational framework for clarifying the type of provenance information capture and the purpose for how it will be used. The existing body of provenance research has grown to a point where the consolidation of design knowledge requires cross-referencing a variety of projects and studies spanning multiple domain areas. The presented framework provides a useful method for organizing and comparing projects and applications designed to support provenance. Such comparison can be especially valuable for the design of evaluations and for connecting the results of previous evaluations. In addition, understanding the types and purposes of provenance can help to advise the design of mechanisms for capturing provenance information, including the appropriate level of granularity.

We do not intend for this paper to be an all-encompassing survey of provenance literature in the field of visualization; rather, we have analyzed the many interpretations of others as a means of creating a framework that encapsulates the core uses of provenance in the field. As with any framework, we acknowledge that the presented organization is not perfect. It will likely need to change in the future, and the distinctions among categories will not always be clean at present. Still, the current organization provides a necessary update to prior frameworks and can serve as a reference point for continued improvement.

Acknowledgments

This work was supported in part by grants from NSF DBI-1260795, IIS-1302755, and MRI-1531491, NIST MSE-20NANB12H181, and DoD USAMRAA-13318046. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, National Institute of Standards and Technology, or Department of Defense. This submission was written by authors acting in their own independent capacity and not on behalf of UT-Battelle, LLC, or its affiliates or successors.

References

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization*, pages 111–117, 2005.

[2] M. K. Anand, S. Bowers, and B. Ludascher. Provenance browser: Displaying and querying scientific workflow provenance graphs. In *IEEE 26th International Conference on Data Engineering (ICDE)*, pages 1201–1204, 2010.

[3] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 55–64, 2010.

[4] R. Ball, C. North, and D. A. Bowman. Move to improve: promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 191–200, 2007.

[5] L. Bavoil, S. P. Callahan, P. J. Crossno, J. Freire, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Vistrails: Enabling interactive multiple-view visualizations. In *IEEE Conference on Visualization*, pages 135–142, 2005.

[6] O. Biton, S. Cohen-Boulakia, S. B. Davidson, and C. S. Hara. Querying and managing provenance through user views in scientific workflows. In *IEEE 24th International Conference on Data Engineering*, pages 1072–1081, 2008.

[7] T. Boren and J. Ramey. Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278, 2000.

[8] M. A. Borkin, C. S. Yeh, M. Boyd, P. Macko, K. Z. Gajos, M. Seltzer, and H. Pfister. Evaluation of filesystem provenance visualization tools. *IEEE*

Transactions on Visualization and Computer Graphics, 19(12):2476–2485, 2013.

[9] E. M. Bowden, M. Jung-Beeman, J. Fleck, and J. Kounios. New approaches to demystifying insight. *Trends in cognitive sciences*, 9(7):322–328, 2005.

[10] U. Braun, A. Shinnar, and M. Seltzer. Securing provenance. In *Proceedings of the 3rd conference on Hot topics in security*, page 4. USENIX Association, 2008.

[11] K. Brodlie, A. Poon, H. Wright, L. Brankin, G. Banecki, and A. Gay. GRASPARC - A problem solving environment integrating computation and visualization. In *Proceedings of IEEE Conference on Visualization (Visualization'93)*, pages 102–109. IEEE, 1993.

[12] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 83–92, 2012.

[13] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. Finding Waldo: Learning about users from their interactions. *IEEE Transactions on visualization and computer graphics*, 20(12), 2014.

[14] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo. VisTrails: visualization meets data management. In *Proceedings of the ACM SIGMOD international conference on Management of data*, pages 745–747, 2006.

[15] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.

[16] P. Chen and B. Plale. Visualizing large scale scientific data provenance. In *SC Companion: High Performance Computing, Networking, Storage and Analysis (SCC)*, pages 1385–1386, 2012.

[17] P. Chen, B. Plale, Y. Cheah, D. Ghoshal, S. Jensen, and Y. Luo. Visualization of network data provenance. In *International Conference on High Performance Computing (HiPC)*, pages 1–9. IEEE, 2012.

[18] H. Chung, S. Yang, N. Massjouni, C. Andrews, R. Kanna, and C. North. VizCept: Supporting synchronous collaboration for constructing visualizations in intelligence analysis. In *IEEE VAST*, pages 107–114, 2010.

[19] G. Closa Santos and J. Mas Pau. A provenance visualization tool for global earth observation system of systems. In *EGU General Assembly Conference Abstracts*, volume 15, page 8266, 2013.

[20] P. Cowley, L. Nowell, and J. Scholtz. Glass box: An instrumented infrastructure for supporting human interaction with information. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 296c–296c, 2005.

[21] S. B. Davidson and J. Freire. Provenance and scientific workflows: challenges and opportunities. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1345–1350, 2008.

[22] N. Del Rio and P. P. Da Silva. Probe-It! visualization support for provenance. In *Advances in Visual Computing*, pages 732–741. Springer, 2007.

[23] B. Demsky. Cross-application data provenance and policy enforcement. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):6, 2011.

[24] M. Derthick and S. F. Roth. Enhancing data exploration with a branching history of user operations. *Knowledge-Based Systems*, 14(1):65–74, 2001.

[25] J. Dobos, N. J. Mitra, and A. Steed. 3D timeline: Reverse engineering of a part-based provenance from consecutive 3D models. In *Computer Graphics Forum*, volume 33, pages 135–144, 2014.

[26] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning process from user interactions. *IEEE Computer Graphics & Applications*, 2009.

[27] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson. GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1663–1672, 2012.

[28] T. Ellkvist, D. Koop, E. W. Anderson, J. Freire, and C. Silva. Using provenance to support real-time collaborative design of workflows. In *Provenance and Annotation of Data and Processes*, pages 266–279. Springer, 2008.

[29] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 473–482, 2012.

[30] A. Endert, C. North, R. Chang, and M. Zhou. Toward usable interactive analytics: Coupling cognition and computation. *Proceedings of the 2014 Workshop on Interactive Data Exploration and Analytics at KDD (IDEA)*, 2014.

[31] K. A. Ericsson and H. A. Simon. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3):178–186, 1998.

[32] J.-D. Fekete, J. J. Van Wijk, J. T. Stasko, and C. North. The value of information visualization. In *Information visualization*, pages 1–18. 2008.

[33] G. A. Fink, C. L. North, A. Endert, and S. Rose. Visualizing cyber security: Usable workspaces. In *6th International Workshop on Visualization for Cyber Security (VizSec)*, pages 45–56, 2009.

[34] J. Freire, D. Koop, E. Santos, and C. T. Silva. Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3):11–21, 2008.

[35] J. Freire, C. T. Silva, S. P. Callahan, E. Santos, C. E. Scheidegger, and H. T. Vo. Managing rapidly-evolving scientific workflows. In *Provenance and Annotation of Data*, pages 10–18. Springer, 2006.

[36] N. Gershon and W. Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001.

[37] D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009.

[38] H. Griethe and H. Schumann. The visualization of uncertain data: Methods and problems. In *SimVis*, pages 143–156, 2006.

[39] D. P. Groth and K. Streefkerk. Provenance and annotation for visual exploration systems. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1500–1510, 2006.

[40] M. Harrower. Representing uncertainty: Does it help people make better decisions. *University Consortium for Geographic Information Science*, 2003.

[41] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196, 2008.

[42] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Queue*, 10(2):30, 2012.

[43] Z. Hensley, J. Sanyal, and J. New. Provenance in sensor data management. *Communications of the ACM*, 57(2):55–62, 2014.

[44] D. A. Holland, M. I. Seltzer, U. Braun, and K.-K. Muniswamy-Reddy. PASSing the provenance challenge. *Concurrency and Computation: Practice and Experience*, 20(5):531–540, 2008.

[45] B. Howe, P. Lawson, R. Bellinger, E. Anderson, E. Santos, J. Freire, C. Scheidegger, A. Baptista, and C. Silva. End-to-end escience: Integrating workflow, query, visualization, and provenance at an ocean observatory. In *IEEE 4th International Conference on eScience*, pages 127–134, 2008.

[46] P. Isenberg and D. Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. In *Computer Graphics Forum*, volume 28, pages 1031–1038, 2009.

[47] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Moller. A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2818–2827, 2013.

[48] R. J. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, 9(2):152–169, 1991.

[49] T. Jankun-Kelly, K.-L. Ma, and M. Gertz. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):357–369, 2007.

[50] W. Javed and N. Elmqvist. ExPlates: spatializing interactive analysis to scaffold visual exploration. In *Proceedings of the Eurographics Conference on Visualization*, pages 441–450, 2013.

[51] N. Kadivar, V. Chen, D. Dunsmuir, E. Lee, C. Qian, J. Dill, C. Shaw, and R. Woodbury. Capturing and supporting the analysis process. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 131–138, 2009.

[52] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann. *Mastering the information age-solving problems with visual analytics*. Florian Mansmann, 2010.

[53] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, (5):44–50, 2013.

[54] D. Kurlander and S. Feiner. Editable graphical histories. In *IEEE Workshop on Visual Languages*, pages 127–134, 1988.

[55] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.

[56] T. Lebo, P. Wang, A. Graves, and D. L. McGuinness. Towards unified provenance granularities. In *Provenance and Annotation of Data and Processes*, pages 39–51. Springer, 2012.

[57] G. Li, A. C. Bragdon, Z. Pan, M. Zhang, S. M. Swartz, D. H. Laidlaw, C. Zhang, H. Liu, and J. Chen. VisBubbles: a workflow-driven framework for scientific data analysis of time-varying biological datasets. In *SIGGRAPH Asia (Posters)*, page 27, 2011.

[58] H. R. Lipford, F. Stukes, W. Dou, M. E. Hawkins, and R. Chang. Helping users recall their reasoning process. In *IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 187–194, 2010.

[59] Y. Liu, O. Sourina, and M. K. Nguyen. Real-time EEG-based human emotion recognition and visualization. In *International Conference on Cyberworlds (CW)*, pages 262–269, 2010.

[60] P. Macko and M. Seltzer. Provenance Map Orbiter: Interactive exploration of large provenance graphs. In *3rd USENIX Workshop on the Theory and Practice of Provenance ((TaPP))*, 2011.

[61] P. Macko and M. Seltzer. A general-purpose provenance library. In *Proceedings of the 4th USENIX conference on Theory and Practice of Provenance*, pages 6–6. USENIX Association, 2012.

[62] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen. Taxonomy-Based Glyph Design with a Case Study on Visualizing Workflows of Biological Experiments. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2603–2612, 2012.

[63] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1633–1642, 2014.

[64] X.-Q. Mai, J. Luo, J.-H. Wu, and Y.-J. Luo. aha! effects in a guessing riddle task: An event-related potential study. *Human brain mapping*, 22(4):261–270, 2004.

[65] J. Matejka, T. Grossman, and G. Fitzmaurice. Patina: Dynamic heatmaps for visualizing application usage. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 3227–3236, 2013.

[66] P. Missier, K. Belhajjame, and J. Cheney. The W3C PROV family of specifications for modelling provenance metadata. In *Proceedings of the 16th International Conference on Extending Database Technology*, pages 773–776. ACM, 2013.

[67] P. Missier, B. Ludascher, S. Bowers, S. Dey, A. Sarkar, B. Shrestha, I. Altintas, M. K. Anand, and C. Goble. Linking multiple workflow provenance traces for interoperable collaborative science. In *Workshop on Workflows in Support of Large-Scale Science (WORKS)*, pages 1–8. IEEE, 2010.

[68] J. Nielsen. *Usability Engineering*. Academic Press, 1993.

[69] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.

[70] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: process+ interaction+ insight. In *ACM CHI Extended Abstracts on Human Factors in Computing Systems*, pages 33–36, 2011.

[71] S. G. Parker and C. R. Johnson. SCIRun: a scientific programming environment for computational steering. In *Proceedings of the 1995 ACM/IEEE conference on Supercomputing*, page 52, 1995.

[72] 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. *Information Visualization*, 8(1):71–84, 2009.

[73] W. A. Pike, J. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.

[74] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pages 109–116. ACM, 2004.

[75] D. J. Pohly, S. McLaughlin, P. McDaniel, and K. Butler. Hi-fi: collecting high-fidelity whole-system provenance. In *Proceedings of the 28th Annual Computer Security Applications Conference*, pages 259–268, 2012.

[76] E. D. Ragan and J. R. Goodall. Evaluation methodology for comparing memory and communication of analytic processes in visual analytics. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 27–34. ACM, 2014.

[77] E. D. Ragan, J. R. Goodall, and A. Tung. Evaluating how level of detail of visual history affects process memory. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 2711–2720. ACM, 2015.

[78] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2356–2365, 2013.

[79] T. Ruotsalo, J. Peltonen, M. Eugster, D. Glowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymki, G. Jacucci, and S. Kaski. Directing exploratory search with interactive intent modeling. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1759–1764, 2013.

[80] S. S. Sahoo, A. Sheth, and C. Henson. Semantic provenance for escience: Managing the deluge of scientific data. *IEEE Internet Computing*, 12(4):46–54, 2008.

[81] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, 2010.

[82] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, 2005.

[83] A. Sarvghad and M. Tory. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the 41st Graphics Interface Conference*, pages 123–130, 2015.

[84] A. Satyanarayan and J. Heer. Authoring narrative visualizations with ellipsis. In *Computer Graphics Forum*, volume 33, pages 361–370. Wiley-Blackwell, 2014.

[85] C. Scheidegger, D. Koop, E. Santos, H. Vo, S. Callahan, J. Freire, and C. Silva. Tackling the provenance challenge one layer at a time. *Concurrency and Computation: Practice and Experience*, 20(5):473–483, 2008.

[86] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[87] S. M. Serra da Cruz, M. L. M. Campos, and M. Mattoso. Towards a taxonomy of provenance in scientific workflow management systems. In *World Conference on Services-I*, pages 259–266, 2009.

[88] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[89] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1237–1246, 2008.

[90] Y. L. Simmhan, B. Plale, D. Gannon, and S. Marru. Performance evaluation of the karma provenance framework for scientific workflows. In *Provenance and Annotation of Data*, pages 222–236. Springer, 2006.

[91] J. Stasko. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 46–53, 2014.

[92] J. J. Thomas and K. A. Cook. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

[93] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the 1st Conference on Visualization*, pages 139–143, 1990.

[94] W. Willett, J. Heer, J. Hellerstein, and M. Agrawala. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 3131–3140, 2011.

[95] K. Wu, J. Chen, W. Pruett, R. L. Hester, et al. Hummod browser: An exploratory visualization tool for the analysis of whole-body physiology simulation data. In *IEEE Symposium on Biological Data Visualization (BioVis)*, pages 97–104, 2013.

[96] K. Xu, S. Attfield, T. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj. Analytic provenance for sensemaking: A research agenda. *IEEE Computer Graphics and Applications*, 35(3):56–64, 2015.

[97] Y. Zhao, M. Wilde, and I. Foster. Applying the virtual data provenance model. In *Provenance and Annotation of Data*, pages 148–161. Springer, 2006.