A Framework for Externalizing Implicit Error Using Visualization

Nina McCurdy, Julie Gerdes, Miriah Meyer

Abstract— This paper presents a framework for externalizing and analyzing expert knowledge about discrepancies in data through the use of visualization. Grounded in an 18-month design study with global health experts, the framework formalizes the notion of data discrepancies as *implicit error*, both in global health data and more broadly. We use the term implicit error to describe measurement error that is inherent to and pervasive throughout a dataset, but that isn't explicitly accounted for or defined. Instead, implicit error exists in the minds of experts, is mainly qualitative, and is accounted for subjectively during expert interpretation of the data. Externalizing knowledge surrounding implicit error can assist in synchronizing, validating, and enhancing interpretation, and can inform error analysis and mitigation. The framework consists of a description of implicit error components that are important for downstream analysis, along with a process model for externalizing and analyzing implicit error using visualization. As a second contribution, we provide a rich, reflective, and verifiable description of our research process as an exemplar summary toward the ongoing inquiry into ways of increasing the validity and transferability of design study research.

Index Terms—implicit error, knowledge externalization, design study

1 Introduction

The research we report on in this paper stems from a six-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 [61]: 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 itself was 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 itself 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 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.

- *N. McCurdy and M. Meyer are with the University of Utah School of Computing, Email: {nina, miriah}@cs.utah.edu.*
- *J. Gerdes is with Texas Tech University, College of Arts & Sciences, E-mail: julie.gerdes@ttu.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

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.

2.1 Public Health Decision Support & 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 [11, 52]. 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 [1, 27, 41, 43, 67]. While this approach is employed in some areas of global health, 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 bio-surveillance [6, 26, 9, 21, 54, 24]. These tools emphasize the important role of situational awareness — the perceptual understanding of the context in which a situation takes place [20] — 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 [27, 40] 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.

2.2 Implicit Error and Knowledge Externalization

The existence of implicit error is well established within public health surveillance [18, 38, 42, 60] and across domains in which humans and societies play a central role in data acquisition, curation, and interpretation [5]. Such domains range from emergency response and disaster operations management [30, 49] to various forms of risk assessment [53], healthcare, and medicine [10]. In public health, surveillance data are often published along with disclaimers like 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* [46]." Attempts to standardize data generation pipelines are a primary approach to minimizing these errors across systems [15]. These efforts are strengthened by methods for evaluating data quality and compliance [47]. There is a general acknowledgement, however, that errors will persist despite these efforts [5, 63]. 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"* [8] — 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."* The work presented in this paper directly attempts to 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 [32, 34, 35, 36].

Visualization is widely recognized as 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 [22, 68]. 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 [12, 22, 72]. 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 [3, 25, 37, 50], and a large body of work within this subfield focuses on visualizing error and uncertainty of geographic data and its associated data attributes [65]. This uncertainty visualization work, however, focuses on quantifiable measures of error and uncertainty, with some attention to categorical measures [19]. 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 [13, 51, 74], for example in areas like human terrain visual analytics [71]. 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.

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. This involves 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. This involves identifying *hotspots* — heavily impacted regions — and *coldspots* — lightly impacted or unaffected regions — and predicting future locations of each. Once these hotspots and coldspots are identified, outbreak data is compared against response data to assess the appropriateness of response efforts.

A key component of outbreak data is *epidemiological surveillance data* (epi-data), which tracks reported cases of a disease and associated health issues through a systematic process of collection, analysis, and dissemination [62]. In the case of the Zika virus, epi-data, which is reported weekly, includes counts of both suspected and confirmed cases of infection along with counts of other related issues like microcephaly and Guillain-Barre Syndrome — the set of reported attributes are referred to as disease *indicators*. Additionally, epi-data is 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 like Dengue, which has been tracked for years. Ento-data was not included in the present study because the focus at the time of tool development was on publicly available case reports. Epi-data is further augmented with geographic data such as rainfall amounts and characterizations of low-lying regions, as well as with demographic datasuch 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 while 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.

4 Process, Artifacts, and Reflection

In this section we report on the core phases of our 18 month-long 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 six 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 [61] 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 to not only build a deployable tool, but 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[1]. 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.

4.1 Learning About Zika Outbreak Analysis

After two 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 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 four 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 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 bi-weekly with our primary collaborators; bi-monthly 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 [33]. 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 D3 sketches, and then proceeded to a high-fidelity visualization tool implemented in D3. 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.

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 highly aware of. Thus, the MOH data, with its finer resolution, promotes 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

[1] http://bit.ly/IEFramework

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. While 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.

4.1.2 Artifact: Data and Tasks Technology Probe

The data and tasks probe developed during this phase, and shown at the top of Figure 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 at 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 to 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 is 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 light-weight 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 experts to support sustainability and continued development of visualization for Zika outbreak analysis.

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

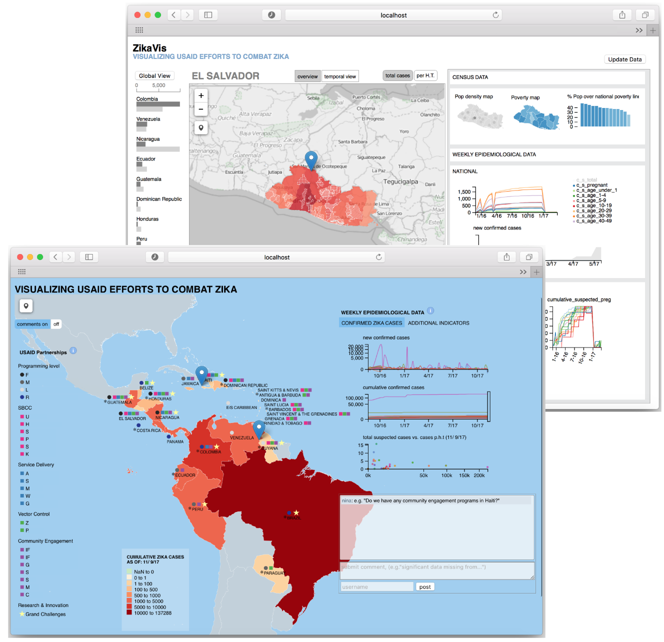

Fig. 1. Technology probes developed during the second (top) and third (bottom) phases of the project. Both versions allow users to explore country and subcountry data from countries' MoH offices (top), and country level data from WHO (bottom). The third phase probe integrates a fully implemented annotation platform.

although the probe was an effective reflection of the data, the data itself was 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's 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.

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 two 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 co-author on this paper.

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 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 presented an opportunity for participants to reiterate thoughts and include new ideas that didn't 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 1.

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%)

Table 1. Results from the workshop poll. The purpose of the poll was to capture initial reactions by the participants about the annotation platform. While 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.

4.2.1 Artifact: Annotation Technology Probe

The annotation technology probe, shown at the bottom of Figure 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.

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.

4.3 Formalization of Learning

The final phase focused on synthesizing and formalizing learning from previous phases [44] surrounding the notion of data discrepancy. This phase took place at the University of Utah and spanned five 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 two of the authors. In the first round, the 54 annotations collected during the annotation workshop were clustered into 3 groups and 4 subgroups. The 3 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 6 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 4.1.1. This larger set of descriptions was then used in a second round of affinity diagramming, conducted by two of the authors. 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."* While 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 [7, 64]. 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 [12, 22, 45, 70, 72] and sociotechnical systems [66], and emergent concepts such as grey literature [48] and systemic bias [2].

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, and describe an instantiation of the framework as a visualization tool for Zika outbreak analysis in Section 6. Our reflective synthesis of this phase is discussed in Section 7.

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.

5.1 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. This is 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 [5]. 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. [4], 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.

5.1.1 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, *grey data*, describes reputed data that is omitted at some stage of the data generation pipeline, due to things such as standardization methods. An example of grey 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 *grey literature* is well established and highly valued within the medical community and refers broadly to findings produced and published outside of traditional academic venues [48]. The analogous grey data is 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 [23, 58].

The third source of implicit error, *retrospective adjustment*, describes downstream updates to previously reported data, resulting in temporal discrepancies. As described in Section 4.1, epi-data is published weekly as static reports and thus, as a consequence, updates and modifications can only be implemented downstream. The regional level WHO data addresses this 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 [39].

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. While magnitude characterizations can be quantitative, 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 supports downstream models for analysis, regardless of whethersuch 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.

5.1.2 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

While useful for analysis and modeling, this characterization 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

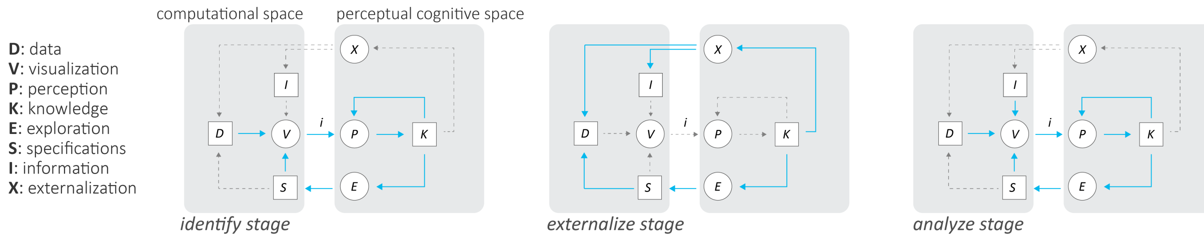

Fig. 2. Process model for externalizing implicit error. The model operates in three stages: In the **identify** stage, insight about the existence of implicit error is generated through the use of a visualization system; in the **externalize** stage, knowledge surrounding implicit error is externalized 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 [68, 12, 72, 22]

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 5.1.1, will enable experts to evaluate reported errors against their own contextual knowledge in order to assess credibility, reliability, and impact [4].

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 to 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 as well. 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 paper 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 [57, 45], as well as adaptations and extensions of these models developed within the visualization community [12, 72, 22].

The DIKW pyramid defines the relationship between data (D), information (I), knowledge (K), and wisdom (W) [17, 16, 70]. *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.

While the DIKW pyramid provides insight into the formal relationship between knowledge, information, and data, 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 [68, 22]. 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 [12, 72]. The models describe a process for externalizing knowledge [72, 22] that incorporates mechanisms for both direct externalization, such as through an annotation interface, and indirect externalization, such as through interaction mining [22]. 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, which incorporates information

The process model is presented in Figure 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 $\boxed{D}$ is visually encoded (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 (P). These processes are both informed by an analyst's knowledge $\boxed{K} \rightarrow$ (P) and yield new knowledge in the form of insights (P) $\rightarrow \boxed{K}$. Here, visualization provides a powerful mechanism for leveraging expert domain knowledge to generate insight about the data [22, 68]. 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 (X). The contextual description about the error is stored as information $\boxed{K} \rightarrow$ (X) $\rightarrow \boxed{I}$, and the characterizing data traits are stored as data $\boxed{K} \rightarrow$ (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: $\boxed{K} \rightarrow$ (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.

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 Figure 3, was built using D3 and Leaflet Javascript libraries and was designed to support the three stages of the process model presented in section 5.2 : 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 reflect 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 is encoded in a choropleth and overlaid on an interactive basemap. The map is linked to a chart-view (Fig. 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 is 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 4.2 and in a large body of literature surrounding the effective use of annotation for narrative and storytelling [55, 59], collaboration and communication [75, 31, 69], externalization of insights [14, 28, 29], and assessments of data quality [4]. Annotations are submitted by dropping markers on regions, countries, or subcountries, which brings up a semistructured annotation template. The template, shown in Figure 3 (right), was designed based on the description of implicit error presented in Section 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 in an online database. To support remote collaboration, new annotations are synchronized via a timer.

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. 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. 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. 3b). Users can toggle between traits through an interactive legend [56], scented with the distribution of categories for each trait [73] (Fig. 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 co-author of this paper. 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 identified several recommendations for others looking to instantiate the implicit error framework:

- Start by designing a visualization system for the existing data — assuming no implicit error — and *then* incorporate annotation mechanisms for externalization.
- Translate the framework constructs into a language that resonates with the domain experts, and employ this language in the annotation mechanism.
- Work with domain experts to generate an example set of annotations for informing the design of annotation encodings.
- Once annotations accumulate, revisit the interface design to support emergent tasks and scalability issues.

We suggest that these recommendations be considered within an iterative, user-centered design process.

7 Discussion

The framework for externalizing implicit error proposed in this paper 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 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. And yet, 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 paper 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

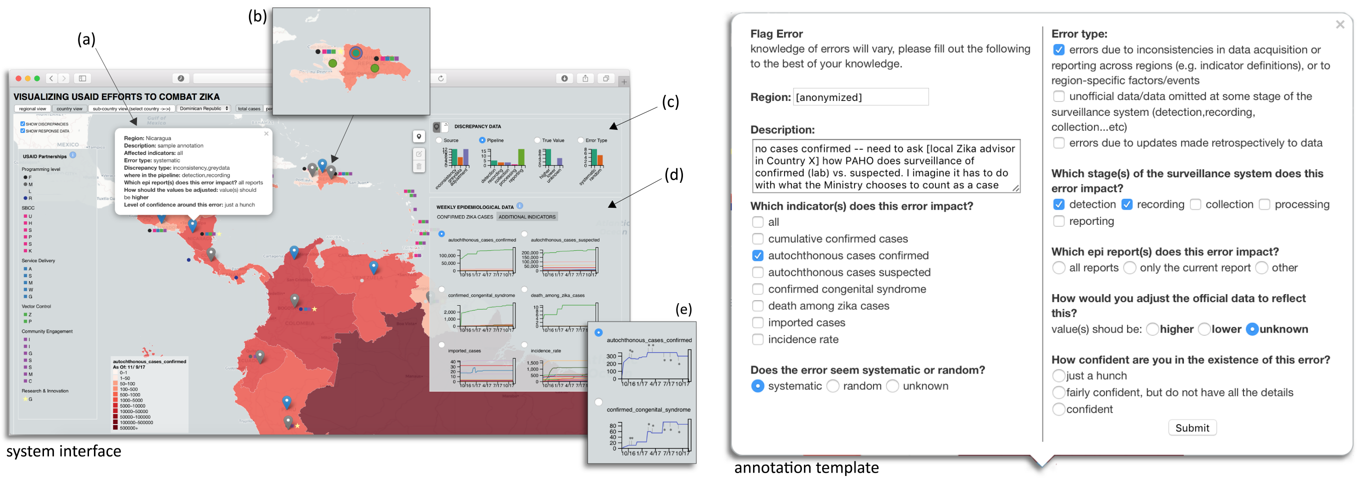

Fig. 3. Prototypical instantiation of the framework for externalizing implicit error. (Right) expert knowledge surrounding implicit error is externalized via an annotation template, shown here featuring an example annotation discussed in section 6. (Left) once submitted, annotations are displayed either (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.

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 [4].

The process model for externalizing and analyzing implicit error described in Section 5.2 modifies existing knowledge-assisted visual analytics models [68, 22] in part by explicitly incorporating the concept of information $\boxed{I}$. 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 inter-organizational relationships. We offer several recommendations based on our experiences reported in this paper. 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. And 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 and this paper 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 on 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.

8 Conclusion And Future Work

We report on an 18-month design study working with Zika experts to investigate challenges surrounding data visualization within the global health community. While we approached this project as a classic design study with the intention of developing a novel visualization system, we instead learned that errors in the data limited the impact of visualization for this community. Through an investigation into these errors, we developed a formalized notion of implicit error and a framework for reasoning about and externalizing implicit error using a visualization system. The framework is grounded in our design study with global health experts, and illustrated through an application to externalizing error in epidemiological surveillance data. This work responds to calls for tools that support capturing and encoding expert knowledge of data quality and uncertainty [4]. We provide an extensive description of our research process and findings, which we tentatively offer as an example of a rich, reflective, and verifiable summary of design study research.

We plan to continue working with our collaborators to further integrate response data and to deploy the system to Zika experts. Additionally, because the Zika virus is spread primarily through mosquito bites, epi-data can be further augmented with ento-data, tracking certain mosquito populations in order to assess underlying transmission risk factors. In future work, we plan to incorporate ento-data and to employ the framework to externalize and analyze implicit error surrounding these data as well. We hope that longer term deployment will enable us to begin to investigate the possibility of modeling characteristics of implicit error as well as the possibility of simulating the mental perceptual adjustments that the global health experts make to incorporate implicit error into their interpretations of data visualizations. More broadly, we plan to explore applications of the framework and of the notion of implicit error to domains outside of global health.

Acknowledgments

We thank our colleagues and collaborators for invaluable discussions regarding this work: Jason Dykes, City, College of London; Mike Kirby, University of Utah; and members of the Visualization Design Lab at the University of Utah. We also thank our collaborators at USAID. This work is supported in part by NSF grant GRF-1747505. The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

REFERENCES

[1] S. Afzal, R. Maciejewski, and D. S. Ebert. Visual analytics decision support environment for epidemic modeling and response evaluation. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 191–200. IEEE, 2011.

[2] M. Alur. Forgotten millions: a case of cultural and systemic bias. *Support for Learning*, 22(4):174–180, 2007.

[3] 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 *Scientific Visualization*, pages 3–27. Springer, 2014.

[4] N. Boukhelifa, M.-E. Perrin, S. Huron, and J. Eagan. How data workers cope with uncertainty: A task characterisation study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3645–3656. ACM, 2017.

[5] G. C. Bowker and S. L. Star. *Sorting things out: Classification and its consequences*. MIT press, 2000.

[6] C. A. Bradley, H. Rolka, D. Walker, and J. Loonsk. Biosense: implementation of a national early event detection and situational awareness system. *MMWR Morb Mortal Wkly Rep*, 54(Suppl):11–19, 2005.

[7] S. Brookfield. Critically reflective practice. *Journal of Continuing Education in the Health Professions*, 18(4), 1998.

[8] J. S. Brown and P. Duguid. *The Social Life of Information: Updated, with a New Preface*. Harvard Business Review Press, 2017.

[9] K. Brown, J. Pavlin, J. Mansfield, E. Elbert, V. Foster, and P. Kelley. Identification and investigation of disease outbreaks by essence. *Journal of Urban Health*, 80:i119–i119, 2003.

[10] P. Carayon and K. E. Wood. Patient safety: The role of human factors and systems engineering. *Studies in Health Technology and Informatics*, 153:23–46, 2010.

[11] 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. *Journal of biomedical informatics*, 51:287–298, 2014.

[12] 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 Computer Graphics and Applications*, 29(1), 2009.

[13] P. Chen, B. Plale, Y.-W. Cheah, D. Ghoshal, S. Jensen, and Y. Luo. Visualization of network data provenance. In *High Performance Computing (HiPC), 2012 19th International Conference on*, pages 1–9. IEEE, 2012.

[14] Y. Chen, S. Barlowe, and J. Yang. Click2annotate: Automated insight externalization with rich semantics. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 155–162. IEEE, 2010.

[15] B. C. Choi. The past, present, and future of public health surveillance. *Scientifica*, 2012, 2012.

[16] D. Clark. Understanding and performance. *Talent & Performance*, 2004.

[17] T. H. Davenport and L. Prusak. *Working knowledge: How organizations manage what they know*. Harvard Business Press, 1998.

[18] S. Declich and A. O. Carter. Public health surveillance: historical origins, methods and evaluation. *Bulletin of the World Health Organization*, 72(2):285, 1994.

[19] L. D. Edwards and E. S. Nelson. Visualizing data certainty: A case study using graduated circle maps. *Cartographic Perspectives*, (38):19–36, 2001.

[20] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.

[21] 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. *Morbidity and Mortality Weekly Report*, pages 32–39, 2004.

[22] 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. 2017.

[23] 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*, 7(12):e1000376, 2010.

[24] 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. *Journal of the American Medical Informatics Association*, 15(2):150–157, 2008.

[25] N. Gershon. Visualization of an imperfect world. *IEEE Computer Graphics and Applications*, 18(4):43–45, 1998.

[26] 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. *Journal of the American Medical Informatics Association*, 19(6):954–959, 2012.

[27] C. Hayes, M. Jilani, A. Yañez, M. Connolly, and J. Duggan. Pandemcap: decision support tool for epidemic management. In *VAHC 2017 (8th workshop on Visual Analytics in Healthcare)-Affiliated with IEEE VIS 2017*. NUI Galway, 2017.

[28] L. Hong and E. H. Chi. Annotate once, appear anywhere: collective foraging for snippets of interest using paragraph fingerprinting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1791–1794. ACM, 2009.

[29] 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 *Proceedings of the working conference on Advanced visual interfaces*, pages 65–72. ACM, 2008.

[30] M. C. Hoyos, R. S. Morales, and R. Akhavan-Tabatabaei. Or models with stochastic components in disaster operations management: A literature survey. *Computers & Industrial Engineering*, 82:183–197, 2015.

[31] J. Hullman, E. Adar, and P. Shah. The impact of social information on visual judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1461–1470. ACM, 2011.

[32] J. Hullman, M. Kay, Y.-S. Kim, and S. Shrestha. Imagining replications: Graphical prediction & discrete visualizations improve recall & estimation of effect uncertainty. *IEEE transactions on visualization and computer graphics*, 24(1):446–456, 2018.

[33] 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 *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24. ACM, 2003.

[34] N. Kijmongkolchai, A. Abdul-Rahman, and M. Chen. Empirically measuring soft knowledge in visualization. In *Computer Graphics Forum*, volume 36, pages 73–85. Wiley Online Library, 2017.

[35] Y.-S. Kim, K. Reinecke, and J. Hullman. Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1375–1386. ACM, 2017.

[36] Y.-S. Kim, K. Reinecke, and J. Hullman. Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation. *IEEE transactions on visualization and computer graphics*, 24(1):760–769, 2018.

[37] C. Kinkeldey, A. M. MacEachren, and J. Schiewe. How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386, 2014.

[38] D. G. Koutsonanos. *Public Health Surveillance Systems for Disease Monitoring, Situational Awareness, and Decision Making Support*. PhD thesis, Emory University, 2014.

[39] T. L. Lash, M. P. Fox, and A. K. Fink. *Applying quantitative bias analysis to epidemiologic data*. Springer Science & Business Media, 2011.

[40] Y. Livnat, J. Agutter, S. Moon, and S. Foresti. Visual correlation for situational awareness. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 95–102. IEEE, 2005.

[41] Y. Livnat, T.-M. Rhyne, and M. Samore. Epinome: A visual-analytics workbench for epidemiology data. *IEEE computer graphics and applications*, 32(2):89–95, 2012.

[42] M. S. Louis. Global health surveillance. *MMWR Surveill Summ*, 61(Suppl.):15–19, 2012.

[43] 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. *Journal of Visual Languages & Computing*, 22(4):268–278, 2011.

[44] N. McCurdy, J. Dykes, and M. Meyer. Action design research and visualization design. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 10–18. ACM, 2016.

[45] I. Nonaka and H. Takeuchi. *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford university press, 1995.

[46] W. H. Organization et al. Global health observatory (gho) data. *URL. Available form: http://www. who. int/gho/tb/en*, 2015.

[47] W. H. Organization et al. Data quality review: a toolkit for facility data

quality assessment: module 1: framework and metrics. 2017.

[48] C. Pappas and I. Williams. Grey literature: its emerging importance. *Journal of Hospital Librarianship*, 11(3):228–234, 2011.

[49] R. W. Perry and M. K. Lindell. Preparedness for emergency response: guidelines for the emergency planning process. *Disasters*, 27(4):336–350, 2003.

[50] K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing*, pages 226–249. Springer, 2012.

[51] 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 transactions on visualization and computer graphics*, 22(1):31–40, 2016.

[52] 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 bio-surveillance and dynamics. In *3rd IEEE Workshop on Visual Text Analytics*, 2013.

[53] J. Rasmussen. Risk management in a dynamic society: a modelling problem. *Safety science*, 27(2-3):183–213, 1997.

[54] 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. *Journal of the American Medical Informatics Association*, 14(5):581–588, 2007.

[55] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe. Chartaccent: Annotation for data-driven storytelling. In *Pacific Visualization Symposium (PacificVis), 2017 IEEE*, pages 230–239. IEEE, 2017.

[56] N. H. Riche, B. Lee, and C. Plaisant. Understanding interactive legends: a comparative evaluation with standard widgets. In *Computer graphics forum*, volume 29, pages 1193–1202. Wiley Online Library, 2010.

[57] J. Rowley. The wisdom hierarchy: representations of the dikw hierarchy. *Journal of information science*, 33(2):163–180, 2007.

[58] 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: Science and Practice*, 3(4):646–659, 2015.

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

[60] T. G. Savel, S. Foldy, C. for Disease Control, Prevention, et al. The role of public health informatics in enhancing public health surveillance. *MMWR Surveill Summ*, 61(Suppl):20–4, 2012.

[61] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2431–2440, 2012.

[62] S. B. Thacker, R. G. Parrish, and F. L. Trowbridge. A method for evaluating systems of epidemiological surveillance. 1988.

[63] S. B. Thacker, J. R. Qualters, L. M. Lee, C. for Disease Control, Prevention, et al. Public health surveillance in the united states: evolution and challenges. *MMWR Surveill Summ*, 61(Suppl):3–9, 2012.

[64] S. Thompson and N. Thompson. *The critically reflective practioner*. Palgrave Macmillan, New York, NY, USA, 2008.

[65] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In *Visualization and Data Analysis 2005*, volume 5669, pages 146–158. International Society for Optics and Photonics, 2005.

[66] E. Trist. The evolution of socio-technical systems. *Occasional paper*, 2:1981, 1981.

[67] 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*, 11(1):37, 2011.

[68] J. J. Van Wijk. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 79–86. IEEE, 2005.

[69] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics*, 13(6), 2007.

[70] K. G. Villholth, E. Lopez-Gunn, K. Conti, A. Garrido, and J. Van Der Gun. *Advances in Groundwater Governance*. CRC Press, 2017.

[71] 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 Transactions on Visualization and Computer Graphics*, 19(12):2139–2148, 2013.

[72] 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. *Computers & Graphics*, 33(5):616–623, 2009.

[73] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.

[74] 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.

[75] 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 transactions on visualization and computer graphics*, 23(1):261–270, 2017.