

Grounding Users in Interpretive Acts: Lessons Learned in the Iterative Design of a Digital Collection

Annie T. Chen, Marium Raza, Yogasai Gazula, Darren Huang, Julianne Mendoza, and Walter G. Andrews



Fig. 1. The Svoboda Diaries Project (<http://www.svobodadiariesproject.org/>)

Abstract— The Svoboda Diaries Project works with personal diaries written at the turn of the 19th century which serve as a rich source on the life, politics, and landscape of Ottoman Iraq. In this paper, we share the path that we have taken to develop a digital resource for different types of users including scholars, researchers, teachers, students and the general public, to embark upon a visual exploration of the diaries. We share this story in two parts: an interview study eliciting feedback from prospective users, and the development of visual and interactive pathways for the collection. In the discussion, we reflect upon the implications of our experience for meeting different user needs and the need to contextualize our work to facilitate interpretation and engender trust. We hope the experiences we share inform the design of other visual platforms for exploring historical digital collections.

Index Terms—visual exploration, timeline, historical documents, digital collection, interview study, interpretation, user research

1 INTRODUCTION

Systems that visualize document collections may focus on various aspects of collections including: metadata, source text, document features and attributes; and concepts, themes, and models based on the documents [1]. In recent years, we have seen increased use of automated textual analysis methods such as topic models [2] and document clustering [1], [3], and visualizations such as clustered displays of documents and word clouds [1], to facilitate exploration and sense-making in document collections. Previous work has also employed visualization to and to support visual reasoning [4], to provide multiple access points, highlight commonalities, and flexible visual pathways to support serendipitous discovery [5]; and to perform visual text analysis, using close and distant reading visualizations, have become increasingly common in digital humanities research [6].

In this paper, we describe our efforts to improve user experience and introduce visual exploration methods to one type of document

collection: a website providing access to digitized historical diaries. There can be particular challenges in the development of digitized historical archives as opposed to digital libraries more broadly, due to the potentially poor preservation state of material and difficulties of dealing with handwritten and illustrated material [7], [8]. Despite recent, digital age advances in studying manuscripts, challenges remain, including diversity of requirements and workflows, a need for greater technical expertise, and the lack of a single tool that can provide needed functionality in one, integrated space [9]. However, by digitizing and providing access to historical artifacts, we can make them available to broader audiences than they would be otherwise. In addition to the data constituted by the object, the metadata associated with an object can be incorporated into visualizations, to afford a richer description of an object's context [10]. Visualizations of historical digital collections can thus offer great affordances for historical research, learning, and discovery.

Issues such as trust, though, can be a concern. There has been interest in factors that are associated with the formation of trust in digital content. For example, previous research has shown that displaying characteristics thought to be relevant to trust (e.g. article stability over time and past editing activity) had an effect on the perceived trustworthiness of information [11].

In this paper, we present a case study of our experience working with the Svoboda Diaries Project, a collection comprised of two sets of historical diaries, through an approach involving stakeholder engagement. We share this story in two parts, first by presenting insights from an interview study of prospective users, and then by

-
- Annie T. Chen, Marium Raza, Yogasai Gazula, Darren Huang, Julianne Mendoza, and Walter G. Andrews are with the University of Washington.
 - Annie T. Chen: atchen@uw.edu.
 - Walter G. Andrews: walter@uw.edu

describing our process for developing visual and interactive pathways to explore the collection. In particular, we consider our lessons learned concerning how contextualization can affect users' willingness and ability to use a collection and its materials, and support the interpretive tasks of historians. In the discussion, we reflect upon how the visual design of the website, its organization, and its interactivity can potentially affect different visitors' utilization of the site, their ability to engage with it effectively for their respective purposes, and their trust in the resource. It is our hope that our experiences can inform the design of other visual and interactive platforms for exploring historical digitized collections.

2 THE SVOBODA DIARIES PROJECT

The Svoboda Diaries Project focuses on the preservation of personal diaries written at the turn of the 19th century, which serve as a rich source on 40 years of the life, politics, and landscape of Ottoman Iraq. The collection is comprised of two main parts, the Joseph Mathia Svoboda diaries, and the travel journals of Alexander Svoboda. Of the former, there are 61 volumes known to be in existence. To date, we have digitized three, and on average, the diaries are 375 pages in length.

There are challenges that we face in digitizing the diaries and making them available for use. There are two major parts of this process: the digitization and publication pipeline, and user-centered design of the website (Figure 2). The first three steps of the digitization pipeline involve digitization of diary pages, transcription, and editing of the transcripts. Edited files are converted into other formats, metadata are assigned, and a diary summary is written. While the majority of the material is in English, there is also material in Arabic and jottings in the margins, which increase complexity.

The design of the website employs a user-centered approach with iterative design cycles involving user needs research, development of wireframes and prototypes; integration with the text, metadata and other content from the digitization and publication pipeline; and evaluation through usability testing. User-centered design and bi-directional communication between the two parts are essential to ensure that the website is usable and approachable to users.

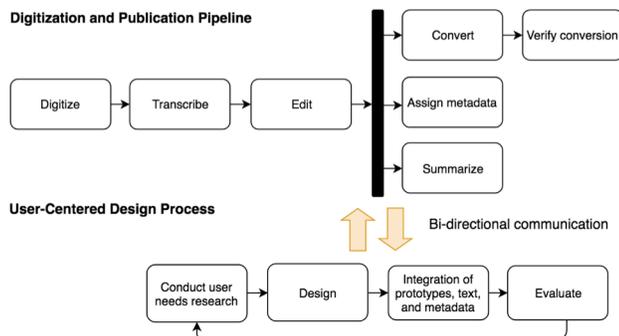


Figure 2. Digitization and Website Production

3 INTERVIEW STUDY

3.1 Methods

Involving stakeholders in projects involving historical research can be important for many reasons. First, different user groups such as historians, librarians, and archivists may have different ways of interacting with materials, and employing a technique such as participatory design can facilitate understanding of how user groups might want to search, browse, and visualize information [12]. Second, involving users in the design of a resource that they will be using can increase their engagement with the resource.

At the outset, we identified three potential types of target users: scholars/researchers, students, and the general public. To better understand the needs of the first group, we identified scholars who possessed research and teaching expertise in related fields such as history and Middle Eastern Studies, as well as directors of labs that engage in digital humanities work. We plan to explore the needs of other user groups in subsequent work.

We conducted contextual and task-based interviews focused on two main parts: 1) the participants' work/research; and 2) participants' impressions of the website. The first part of the interview was comprised of questions to help us better understand participants' research and teaching interests, and their work routines. In the second part of the interview, we introduced participants to the website and asked them to comment on the design, in particular on the clarity of presentation, usefulness, and ease of use.

We conducted two rounds of interviews, one set prior to the launch of our new website, and one set after. Depending on when they were interviewed, participants were asked either to view the website prior to or after the re-design. In the interviews that were conducted after the launch of the re-designed website, we also asked participants to perform various tasks with aspects of the re-designed website such as the new feature for visual exploration, the event timeline, and a diary viewer, which also included a visual component, while 'thinking aloud.' Use of the think-aloud protocol is a common task-based usability testing technique [13].

We transcribed the interviews verbatim and then engaged in qualitative analysis of the interviews. We employed the Grounded Theory Method (GTM) to code the data [14]. In grounded theory, researchers code data for themes and concepts of interest and then engage in an iterative process of comparing and revising codes, also called constant comparison, to solidify the codes [15]. Two authors (AC and YG) independently coded the transcripts and then resolved disagreements by consensus. We report findings based on the most salient codes that appeared in our data.

3.2 Results

We conducted 6 interviews prior to the website re-design and 3 interviews after the re-design. The participants had differing levels of familiarity with the subject matter of the diaries, historical research, and digital humanities research, as well as different degrees of experience in their fields. The interview structure was helpful to characterize the nature and diversity of participants' work routines, which could affect their website interactions. For example, there was variability in terms of whether individuals worked with primarily physical materials, digitized materials, or both.

The interviewees' work involved various decisions that could be influenced by, or subsequently affect, others' ability to perform interpretive tasks. For example, P1 talked about the decision to render a poem in meter or prose during transcription, and P4 talked about the decision to note changes to a manuscript on one version, or to have separate versions. These topics came up in the discussion of participants' daily work, but also served as reminders to us of the need for accountability in the interpretive acts in which we engaged.

With regard to interviewees' interests in the website, some participants were more interested in the project history, while others expressed a desire to have more information contextualizing the materials provided. Participants expressed appreciation of the site making primary documents accessible through the original diary images, which were displayed side-by-side with transcriptions, and the utility of this for pedagogy and research. Participants also spoke of the difficulty of reading the handwriting, and the utility of having a resource such as this, which could alleviate this difficulty for students and researchers alike, while at the same time facilitating double-checking of the original text as needed.

One prominent theme was that there were gaps in information and areas of confusion. In particular, there was a need for high-level information about the collection, including, at the most basic level, what the collection was comprised of and the extent of it: "it might

just be good to have an overview of exactly what's here, how many pages, how much of it has been digitized, how much has been transcribed, you know, plans to keep going”(P9). There were aspects of the visual and textual organization that were confusing to users. For example, the previous website included icons leading to other lab projects at the center of the index page, which distracted users and led to confusion about the project. As P5 put it, “the visual cue doesn't match the textual cue.”

There was a need for more explanatory text to provide context such as who the individuals mentioned in the diaries were, and instructions on how to use the website features and tools. Interviewees also expressed the need to explain the file formats offered and which users might find certain formats useful: “it's great that you have all these different formats, you need to contextualize these different formats. And maybe even say like ‘This format is best for this type of user’... XML is better for people who want to do textual analysis, PDF is going to be better for those who want to read it. HTML is better for people who want to use screen readers...”(P5).

Participants also wanted to be able to search through the diaries. Potential searches of interest included who visited specific places and when quarantine periods began and ended, reflecting the potential for the diaries to be used to address different research questions. Other desired functionality included the ability to focus on specific years and to account for spelling variants. For example, one participant noted: “the way he's spelling the names of Iraqi places and people in the 19th century in his own way is not something that I think the average researcher would know off the top of their heads. So for instance the way Diala is spelled here is not usually the way you'd see it in English today”(P2).

To summarize, there were several main takeaways. First, there was a need for a clearer, high-level introduction to the collection and how to use it, and to tailor content to different types of users. Second, there was a need to add context and provide visual cues to help orient users. Third, there was a need to enable users to interact more with the sources, in ways that suited their interests. Thus, the interview results demonstrated the need for increased visual and textual clarity, modes of interaction suiting particular user needs, and increased support for discovery. We have implemented these takeaways to varying degrees in the current version of the website.

4 DEVELOPING VISUAL AND INTERACTIVE PATHWAYS

In the re-design of the website, we set out to create multiple visual pathways to explore the collection. There were several motivations to do so, including: providing a more interactive way for people to explore, enabling viewers to identify important historical events, and facilitating discovery of patterns in the diaries over time. We also sought to facilitate a more kinesthetic experience for users to interact with and consider the diaries.

The first of these was to develop a visual timeline for exploring

events that were mentioned in the diaries (Figure 3, left). This timeline offers users multiple mechanisms for exploration: 1) clicking backwards and forwards through events depicted in a central pane; 2) studying the events in the timeline; and 3) exploring events assigned to categories of interest (e.g. Religious/Cultural, Economic). The center pane depicts a featured event, usually with an image, event name, description, and link to the source diary. Clicking on the diary link brings up images of the diary and textual transcription side-by-side (Figure 3, right).

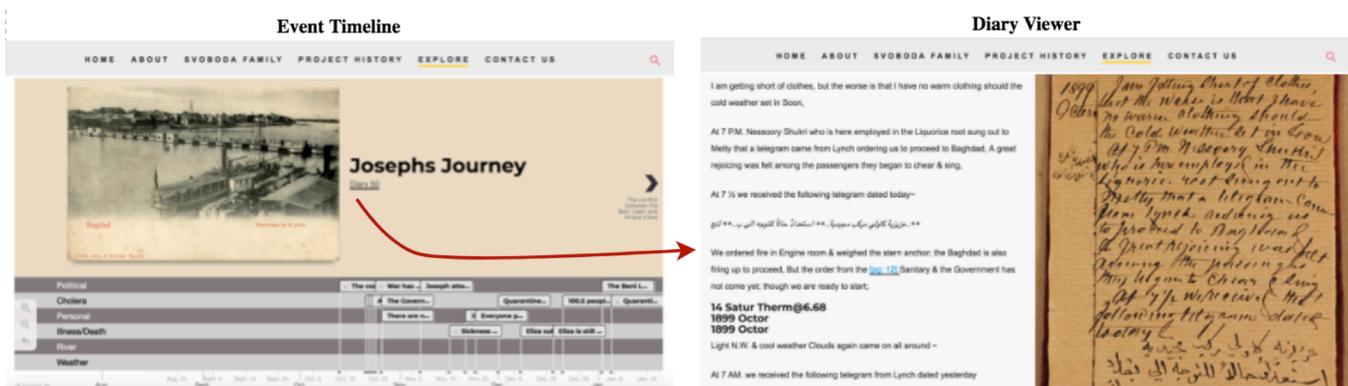
The timeline employs Timeline JS, an open-source tool for building visual timelines (<https://timeline.knightlab.com/>). Even in the use of existing, open-source packages, there can be difficulties in adapting packages to fit project needs. In this case, one particular challenge lies in the complexity of the information to be visualized. We describe some of the challenges that we faced and continue to face as we develop this representation, which is critical to effective display of the material and subsequent exploration by the user.

The visual timeline requires event metadata, which is produced as a part of the digitization pipeline (Fig. 2). We employ a manual process in which events of interest are identified and classified by the project team. There were various difficulties that we faced, including the development of a common set of event categories that work with all of the diaries. Initially, Social, Political, Economic, Religious/Cultural, and Disease/Illness were chosen to represent events. In practice, these were too ambiguous to be used in categorization. Another point of ambiguity was the selection of which events to provide metadata for. We selected events based on anticipated importance to future users: notes on water level or cargo transported, for example, were neglected, while meetings with dignitaries, cholera epidemics, and religious holidays were included. Another challenge was the need of prospective users to differentiate between personal events in the lives of the persons mentioned in the diaries, and events of historical significance. Lastly, there were seemingly significant events that did not fit into the categories.

Currently, the process by which the collections are read, events extracted, and then categorized, is documented to ensure standardization. Categories are periodically re-evaluated by the team to determine whether or not they effectively describe and characterize diary events. The current categories include Family/Personal, Local conflict, National/International event, Ship/Voyage, and Weather/Natural disaster, Epidemic, and Personal Illness. Example events and associated metadata appear in Table 2.

Table 2. Event Metadata Schema and Examples from Diary 49

Page	Date	Headline	Category
275	6/2/1899	Joseph's brother in law is sick	Family/Personal
280	6/5/1899	Cholera detected in Baghdad	Disease/Illness
282	6/6/1899	Tribal unrest outside Basreh	Local conflict



Users can explore by: 1) clicking on arrows to advance the events; 2) browsing through timeline entries; and 3) clicking on events to bring them into focus.

Clicking on the link for a specific event brings up the source diary in the diary viewer. Transcriptions and diary images appear side by side, for perusal by the user.

Figure 3. Explore Section: Event Timeline and Diary Viewer Interactive Pathway

After the website was redesigned, we incorporated tasks into the interviews to obtain feedback on prospective users' reactions to the timeline and diary viewer. Overall, participants appreciated the interactivity of the website. The diary viewer afforded visitors the opportunity to double-check the transcriptions against the actual diary images. One participant reflected upon the difficulty of categorizing events on the timeline: "it seems like... it's really hard to tell to distinguish a death from something that's personal for him happening. I know sometimes he does talk about people he doesn't know death. It seems like some of these deaths would be within the personal... I'm not sure what your criteria is for that"(P7). Thus, we also saw the themes of primary document accessibility and the need for more explanation of work process within the Explore section.

5 DISCUSSION

5.1 Endeavoring to Create an Interactive and Visual Experience

As we move towards creating a more engaging experience, we reflect on challenges that lie ahead. The task of preservation can be a difficult one, as the materiality and texture of manuscripts [16], the depth of original audio recordings [17], and other aspects of original material can be important yet difficult to capture in digital form. In this case study, we presented our approach to preserve and improve the user's experience of a digital manuscript collection. As part of a user-centered design approach, we have been conducting interviews with prospective users to better understand their needs, and we have been reminded of the imperative to coordinate visual and textual cues to facilitate interaction and sense-making. Alignment of visual and textual information can serve to ground users, and the absence of it can lead to confusion and disengagement. Moving forward, a critical challenge we face is ensuring that the metadata we develop supports the timeline, diary viewing, and search functionalities the site offers.

Though interviewees appreciated the greater interactivity of the website, they also advocated for even greater support for interactivity and flexibility. For example, P7 suggested being able to pull up multiple diary images and/or transcriptions side-by-side and being able to minimize elements at will, to focus on other elements.

5.2 Implications for Different User Groups

Our work on expanding the visual and interactive experience of the Svoboda Diaries has resulted in greater understanding of the needs of different types of users and the ways in which they might use the website. In Table 3, we consider the implications of different aspects of the website for different user groups. At the outset, there were aspects of the site that generally provoked confusion, which could be mitigated by reorganizing the site visually, providing more contextual information, and including explanatory text so users could understand how the content was relevant or useful to them.

With regard to research, we learned of the desire for more sophisticated search and text analysis capabilities. Those who were actively engaged in digital humanities and were familiar with automated textual analysis tended to comment more on the need for a concise description of the corpus and the project and a bulk download functionality, whereas those who were accustomed to working with primary documents emphasized search functionality.

Many interviewees commented on the potential use of the diaries in pedagogy. P6 remarked he could use the travel journals to expose students to primary sources and when teaching Arabic. P3 suggested that we include exercises, such as having students read the diaries and consider them along with contextualizing information created by the team: "Instead of saying that it's a summary say it's an interpretation or narration of events... It'd be really interesting to get students to read the diary to pair it with that initial interpretation that... someone on your team has done and have the students interpret it... with the assessment of the primary source."

Table 3. Recommendations for Different User Groups

User Group	Facet	Recommendation
General	Clarity of context	Provide information about how the parts of the project fit together
	Clarity of instructions	Tell users how they can use the tools and materials provided
	Appeal	Tell users why the content is interesting
Research	Personalization	Make it apparent to different types of users how the content applies
	Searchability	Provide alternative ways to search the diaries
Pedagogical	Interactivity	Support multiple ways of interacting with the text
	Accessibility of primary sources	<ul style="list-style-type: none"> • Provide contextual information about persons and events in the diaries • Provide sample activities

5.3 Implications for Engendering Trust with Historical Document Collections

The process of making historical artifacts available and the process of working with digitized artifacts can both involve interpretive acts. In the process of making the diaries available, we make decisions regarding digitization, transcription, and metadata assignment. Team members transcribe the original primary source materials to make them accessible and create additional materials, such as summaries, to aid in the process of learning more about the diaries. However, editing transcriptions for unclear words also introduces potential differences in details that might be interpreted differently by different transcribers. Though these instances are usually resolved by consensus, they nevertheless involve making judgments. In addition, the creation of summaries by individual transcribers introduces an additional layer of interpretation.

Sternfeld writes that the act of relating units of historical information through a process of selection, search, and the application of metadata reflects an act of historical interpretation, and contextualization of this act contributes to trustworthiness [19]. As P3's remark illustrated, the work that we do to provide descriptions of diaries makes them more approachable and easier to understand, but they are also interpretive acts. The decisions that we make concerning how to make the diaries available, which events to include, and what to include in the summaries, are all decisions that affect subsequent interactions with the material.

In our study, participants expressed confusion due to the lack of clear explanations of the digital publication process. Providing information about the processes and standards used could serve to not only facilitate subsequent interpretation, but also nurture trust. Increased documentation and transparency is a goal we will continue to work towards in future development work.

6 CONCLUSION

In this paper, we presented a case study involving iterative design of a historical digital collection, the Svoboda Diaries. We focus on two parts of the process, an interview study to elicit prospective user feedback, and the development of visual and interactive pathways to facilitate exploration. The interviews identified aspects of the website that participants found confusing and undermined engagement. Participants appreciated the visual and interactive elements of the redesigned website, but expressed the need for more information underlying conceptual elements such as timeline events. Greater contextualization and interactivity were identified as design priorities. In future work, we plan to also conduct user research with other target user groups, such as the general public and students.

ACKNOWLEDGMENTS

The authors wish to thank interview study participants for their feedback, and the interns, particularly, Corina Geier and Daniel Kim, for their contributions to the user research and website re-design.

REFERENCES

- [1] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko, "Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw," *IEEE Trans Vis Comput Graph*, vol. 19, no. 10, pp. 1646–1663, Oct. 2013.
- [2] J. N. Vilaplana and M. Pérez-Montoro, "Diggersdiaries: Using text analysis to support exploration and reading in a large document collection," *Unpublished*, 2017.
- [3] M. Brehmer, S. Ingram, J. Stray, and T. Munzner, "Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool for Investigative Journalists," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2271–2280, Dec. 2014.
- [4] N. Kodagoda, S. Attfield, B. L. W. Wong, C. Rooney, and S. Choudhury, "Using Interactive Visual Reasoning to Support Sense-Making: Implications for Design," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2217–2226, Dec. 2013.
- [5] A. Thudt, U. Hinrichs, and S. Carpendale, "The bohemian bookshelf: supporting serendipitous book discoveries through information visualization," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, Austin, Texas, USA, 2012, p. 1461.
- [6] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann, "Visual Text Analysis in Digital Humanities: Visual Text Analysis in Digital Humanities," *Computer Graphics Forum*, vol. 36, no. 6, pp. 226–250, Sep. 2017.
- [7] A. Katifori, E. Torou, C. Vassilakis, and C. Halatsis, "Supporting Research in Historical Archives: Historical Information Visualization and Modeling Requirements," in *2008 12th International Conference Information Visualisation*, 2008, pp. 32–37.
- [8] A. Weber, M. Ameryan, K. Wolstencroft, L. Stork, M. Heerlien, and L. Schomaker, "Towards a Digital Infrastructure for Illustrated Handwritten Archives," in *Digital Cultural Heritage: Final Conference of the Marie Skłodowska-Curie Initial Training Network for Digital Cultural Heritage, ITN-DCH 2017, Olimje, Slovenia, May 23–25, 2017, Revised Selected Papers*, M. Ioannides, Ed. Cham: Springer International Publishing, 2018, pp. 155–166.
- [9] B. Almas, E. Khazraee, M. T. Miller, and J. Westgard, "Manuscript Study in Digital Spaces: The State of the Field and New Ways Forward," *DHQ*, vol. 012, no. 2, Jul. 2018.
- [10] F. Windhager *et al.*, "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 6, pp. 2311–2330, Jun. 2019.
- [11] A. Kittur, B. Suh, and E. H. Chi, "Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia," in *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, San Diego, CA, USA, 2008, p. 477.
- [12] E. Dimara, A. Bezerianos, and P. Dragicevic, "Narratives in Crowdsourced Evaluation of Visualizations: A Double-Edged Sword?," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2017, pp. 5475–5484.
- [13] M. W. M. Jaspers, "A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence," *International Journal of Medical Informatics*, vol. 78, no. 5, pp. 340–353, May 2009.
- [14] B. G. Glaser and A. L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Pub. Co., 1967.
- [15] K. Charmaz, "'Discovering' Chronic Illness: Using Grounded Theory," *Social Science & Medicine*, vol. 30, no. 11, pp. 116–1172, 1990.
- [16] E. Treharne, "Fleshing out the text: The transcendent manuscript in the digital age," *postmedieval: a journal of medieval cultural studies*, vol. 4, no. 4, pp. 465–478, Dec. 2013.
- [17] J. Rosen, "The Day the Music Burned," *The New York Times*, 11-Jun-2019.
- [18] N. Eichenberger, C. H. Johnson, and L. Seige, "A week of manuscript studies with the digital workspace 'mirador@ubleipzig,'" *Informationspraxis*, vol. Bd. 3, p. Nr. 2 (2017)-, Jan. 2018.
- [19] J. Sternfeld, "Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization," *The American Archivist*, vol. 74, no. 2, pp. 544–575, Sep. 2011.