EG UK Computer Graphics & Visual Computing (2024)
A. Slingsby and D. Hunter (Editors)

Reflections on the Evolution of the BookTracker Visualization Platform

Y. Xing[1], C. Dondi[2], R. Borgo[1], and A. Abdul-Rahman[1]

[1]King's College London, United Kingdom
[2]Sapienza University of Rome, Italy

Abstract

Understanding the trade data of historical books is crucial for researchers investigating the distribution and provenance of Incunabula (books printed between 1450 and 1500). We incrementally developed BookTracker, a platform featuring multiple visualization and visual analytics applications to support these research efforts. This platform leverages data from the Material Evidence in Incunabula (MEI) database, which meticulously records detailed information on the provenance, ownership, and use of 15th-century printed books. BookTracker began with a focus on providing visualization and visual analytical solutions to effectively present each book provenance's chronological and geographical information. Through three years of collaborative work with domain experts, we continually explored the Material Evidence in Incunabula (MEI) data and discovered more possibilities for visualization to represent this rich information. Gradually, a suite of specialized visualization tools for specific analytical purposes was developed, including DanteSearchVis, DanteExploreVis, KURF2022, KURF2023, and OwnershipTracker. These tools now comprise the BookTracker platform, which has evolved to explore various features and aspects of the data. This paper details the evolution of BookTracker's design and development alongside domain experts, highlighting the reflections and lessons learned from its application in various research projects. We discuss this long-term collaborative visualization project, hoping to offer our experience as a case study for similar research in the future.

CCS Concepts

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

1. Introduction

Books have historically circulated widely, disseminating knowledge, ideas, and information while changing hands. The Material Evidence in Incunabula (MEI) database [CER15] has been compiling data on the ownership and use of the earliest printed books known as Incunabula. This comprehensive dataset, enriched with material evidence (inscriptions, annotations, decoration, and binding styles), documentary evidence (historical library catalogs, acquisition lists), and bibliographical evidence (auction and booksellers catalogs), allows researchers to trace the circulation and provenance of these books across Europe and the USA over several centuries. Visualizing the movement of historical books is both fascinating and challenging. In a time predating modern logistics services like FedEx or Amazon, thousands of books made their way to their current locations across multiple continents over centuries. These books carry records of their journeys through material evidence such as manuscript annotations, decorations, and bindings. Capturing these movements, which encompass both spatial and temporal dimensions, can answer numerous research questions for historians. While tracking the movement of books in the digital age might appear straightforward, doing so for historical books involves significantly greater complexity.

Book historians face considerable challenges in analyzing book trade data due to its complexity and volume. Nonetheless, such data is indispensable for reconstructing historical narratives and validating scholarly hypotheses. For example, understanding the journey of a single book through various owners and locations over centuries necessitates detailed records provided by the MEI database, alongside effective visualization tools to interpret these records. To tackle this challenge, we have collaborated closely with experts in historical book research, notably including the founder of the MEI database. This collaboration has given rise to a long-term visualization project named BookTracker, accompanied by an online platform of the same name. Under the BookTracker initiative, we have progressively undertaken a series of visualization design study projects to explore visualization solutions to present and elucidate the movement of books from diverse perspectives. BookTracker integrates five tools–DanteSearchVis, DanteExploreVis, KURF2022, KURF2023, and OwnershipTracker–each designed for specific analytical purposes. Developed over the last three years, these tools reveal patterns and connections within the data, such as the circulation and distribution of single and multiple book copies, the dispersal of books from religious orders, and the dynamics of book ownership.

DOI: 10.2312/cgvc.20241234

The success of BookTracker is rooted in ongoing collaboration between visualization researchers and book historians. Essential factors include the expertise of domain specialists, easy access to data through API, regular feedback sessions, and the involvement of undergraduate students in summer projects for tool development. This paper details the BookTracker platform's evolution, discussing each tool's development and integration into a cohesive research resource. We also reflect on the project's conduct, the importance of sustained collaboration, and insights gained from the design and evaluation processes. We hope to offer our experience as a case study for similar research.

2. Related Work

Before delving into the historical book domain and the development of BookTracker, let us examine current visualization techniques for digital humanities and similar collaborative projects.

2.1. Visualization for Humanities

Over the last decade, various scholars have emphasized the need for humanities to develop visualizations explicitly tailored to the unique requirements of humanities projects [BSS20, BEAC*18]. Drucker argued that visualization techniques from the natural sciences do not meet humanities scholars' needs because they often simplify research data and assume knowledge to be observer-independent and factual. Unlike scientific data, which is considered given, humanities data should be interpreted as selected and constructed. This distinction implies that data visualizations in the humanities should always provide contextualization through multiple layers of information, presenting a significant challenge for existing visualization techniques [Dru11].

Integrating close and distant reading is one promising area bridging the gap between visualization and humanities. Close reading involves detailed inspection of individual texts, while distant reading analyzes large text collections. Historically, visualization has been more tied to distant reading, but a growing trend combines the two. Projects like Serendip [AKV*14] and VarifocalReader [KJW*14] connect document cluster views to close reading, building trust in statistical text models. Tools such as Poemage [MLCM16] further enable new interactions at the close reading level.

Another fruitful area is geospatial data and mapping. Maps and timelines situate information in space and time, uncovering relationships across boundaries. Projects such as the Linguistic Landscapes of Beirut (LLB) explore Lebanon's multilingual public language through digital maps [Wri20], and Visualizing Medieval Places connects place names in medieval French texts using spatio-temporal visualization [Wri17].

Text similarity detection is crucial in large-scale text mining and historical document analysis. Visual analytics have been applied to identify variance in medieval poetry [JW17] and commonplaces in 18^{th}-century literature [ARRO*17], providing new insights.

Emerging challenges also include serendipitous discovery and the evolving role of aesthetics in visualization. Inspired by library sciences, exploratory visualization highlights the importance of visually pleasing presentations alongside functional roles [THC12].

2.2. Collaborative Visualization Projects

Collaborative visualization projects are redefining research areas by integrating humanities and visualization to advance semi-automatic argumentation analysis [EAHJG*17]. The visualization research process itself holds significant value beyond the final tools. These efforts demonstrate the potential for stronger, more integrated collaborations between visualization and humanities, driving innovation and expanding the impact of both fields.

Examples of such collaborations include The Angkor Project [CMW*17] and The Quill Project [CARM17], highlighting the power of sustained interdisciplinary efforts. The Quill Project focuses on the history of constitutional conventions, providing tools to model, analyze, and visualize the complex processes of drafting and negotiating constitutional texts. This project brings together historians, legal scholars, and computer scientists to create a comprehensive digital resource that enhances understanding and facilitates research on constitutional history. Similarly, The Angkor Project aims to map and analyze the extensive archaeological site of Angkor in Cambodia. By integrating geographic information systems (GIS), remote sensing, and historical research, the project provides detailed visualizations of the site's structure, evolution, and historical significance. This collaboration advances archaeological methods and contributes to the preservation and understanding of cultural heritage. These examples highlight the value of long-term collaborations in addressing complex research questions and developing innovative solutions that benefit both the humanities and visualization communities.

3. Domain Background and Data

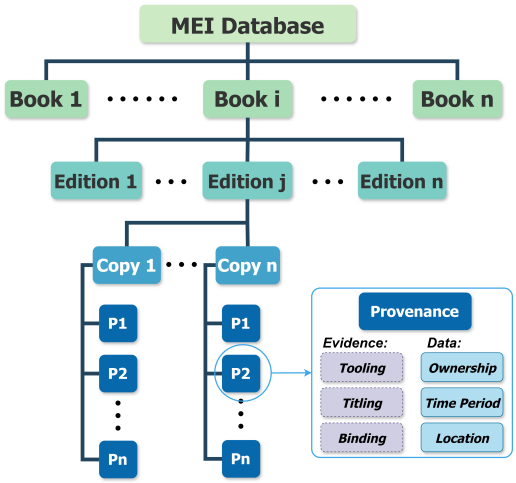


Figure 1: *This figure illustrates the hierarchical structure of the MEI database, highlighting the material evidence and digital records contained within each provenance block.*

The evolution of BookTracker is deeply rooted in the groundwork laid by the 15cBOOKTRADE Project [Don13, Don24, Don20] and the Material Evidence in Incunabula(MEI) Database [CER15].

3.1. The 15cBOOKTRADE Project

The ERC-funded 15cBOOKTRADE Project [Don13, Don24, Don20] aimed to assess the impact of the first 50 years of the European printing revolution (1450-1500) on European society, mainly focusing on the transition from the medieval to the early modern period. By utilizing material evidence from thousands of surviving books, the project addressed fundamental questions about the introduction of printing in the West, which had previously eluded scholars due to a lack of evidence and effective tools for analyzing existing data.

The project employs 15^{th}-century printed books as historical artifacts, with half a million surviving examples preserved in approximately 4000 libraries, mainly in Europe and the United States. Scholars specializing in historical books collect various types of data from each book copy–including inscriptions from previous owners, marginal annotations, decorative and binding styles, pricing details, and stamps. They also examine documentary records, such as historical library catalogs, acquisition lists, and bibliographical sources from auctions and bookseller catalogs. All gathered information is meticulously recorded in a database to trace the provenance and document the history of each book from its original publication place and time to its current repository.

3.2. Material Evidence in Incunabula Database

The Material Evidence in Incunabula (MEI) database was developed to provide a substantive representation of the circulation of book copies throughout the centuries, from their place of production to their present locations. It is freely accessible online and was significantly developed during the 15cBOOKTRADE Project to support the extensive data collection of up to half a million records.

The MEI database documents the distribution, sale, acquisition, and use of thousands of surviving 15^{th}-century printed book copies, encompassing dozens of fields related to provenance and acquisition methods. It records former owners, defined by gender, status (lay, religious), and profession, and includes images from relevant incunabula. One of the most impressive aspects of MEI is that each book copy contains a series of provenance blocks that capture its life journey. Each piece of provenance evidence is recorded in a separate block, tagged geographically and chronologically, and visually represented in the MEI records as a sequence of provenance blocks. All book copies are the basic units of a hierarchical structure, with upper layers of editions and different books. The hierarchy structure, evidence, and resulting digital records for each copy's provenance are illustrated in Fig. 1.

3.3. Previous Visualization Tools

Before the development of BookTracker, few visualization tools were dedicated to exploring and explaining MEI data. The ATLAS OF EARLY PRINTING [PHS24] is one such software, using GIS maps of Europe to depict the spread of printing, trade routes, and bishoprics. However, domain experts believe this software could benefit from an upgrade in its functionality.

The 15cBOOKTRADE Project launched 15cVis (Fig. 2 a), an innovative software designed to visualize the mobility of books using MEI data. This tool traced the journey of books from their origins to their present locations, employing Cartesian diagrams to depict geographical locations on the vertical axis and a chronological timeline from 1450 to the present on the horizontal axis. Although colors, shapes, and positions were utilized to encode geographical information, this approach inadvertently masked some finer geographical details. Despite its sophisticated features, domain experts expressed concerns when discussing their experiences with the tool, particularly highlighting the significant uncertainties in the temporal attributes of the MEI data. These uncertainties significantly constrained the tool's effectiveness in displaying time-based visualizations. Moreover, the experts indicated that a precise chronological representation did not substantially benefit their research. Instead, they recommended a greater focus on the spatial transitions of books, suggesting a reduction in temporal detail to enhance the emphasis on geographical mobility. This shift toward more intuitive map-based visual representations prompted historians to seek new collaborative opportunities in visualization. Additionally, the cessation of support and maintenance for 15cVis created a void that the BookTracker project aimed to fill, driven by the previous tool's unavailability.

3.4. Transition to BookTracker

Following the 15cBOOKTRADE Project, the need for more sophisticated visualization tools became evident, leading to the development of BookTracker. This cross-disciplinary project between visualization experts and humanities scholars aimed to further explore and utilize the rich data from the MEI database. The goal was to create tools that effectively visualize book provenance data and address various research questions from different perspectives.

4. The Evolution of BookTracker

In this section, we will discuss BookTracker's development journey, the motivations behind its continuous evolution, its key components, the evaluation and testing methods employed, and the integration of these components into a cohesive platform.

4.1. Development Timeline

The development of BookTracker's projects spanned several years, reflecting a gradual expansion in focus and capabilities. Each project had specific objectives, resulting in different visualization solutions that built on previous experiences and addressed new research challenges.

DanteSearchVis & DanteExploreVis (2021-2022): The initial development phase of BookTracker saw the creation of DanteSearchVis and DanteExploreVis (Fig 2 b&c) by a PhD student specializing in visualization. These tools were designed to analyze a sub-dataset from MEI, specifically focusing on the surviving book copies of the first illustrated edition of Dante's Commedia, printed in 1481.DanteSearchVis enabled precise exploration of user-defined book movement paths through query-based searches and visualization techniques. DanteExploreVis facilitated the comprehensive exploration and presentation of temporal and geographical book trade data from multiple perspectives.

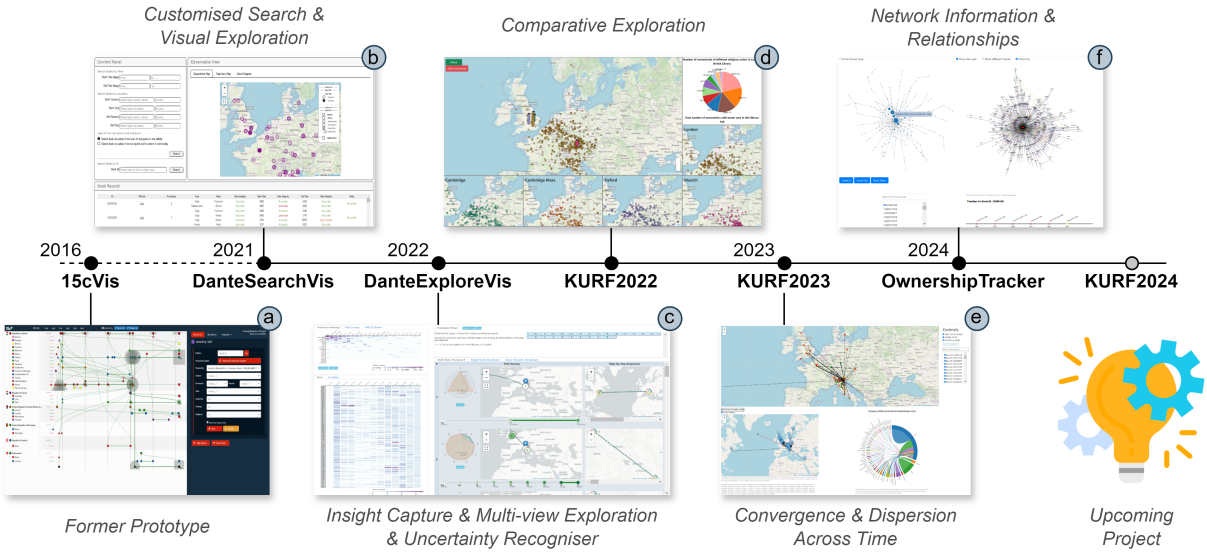


Figure 2: *This figure illustrates the development timeline for each component (b, c, d, e, f) of the BookTracker platform, highlighting the main domain problems they aim to solve. Additionally, a previous visualization attempt is shown in a.*

KURF2022 (2022): Following the success of previous projects, the domain expert proposed a new research focus of exploring the dispersal of books from religious orders throughout Europe during the sixteenth century and beyond. This led to the development of KURF2022 (Fig 2 d), a project driven by undergraduate students in computer science who were also interested in this topic. These students contributed to the engineering aspects of the visualization tool as part of their summer 2-month research internship. The resulting tool provided specialized features for tracking the distribution patterns of these books, offering insights into the survival of monastic book culture.

KURF2023 (2023): Building on the previous collaboration experience with undergraduate students, KURF2023 (Fig 2 e) was launched, with new student developers joining. This project focused on visualizing the phenomenon of book gathering and dispersal due to significant historical events, specifically analyzing the book trade data of the Venice Benedictines' library, S. Giorgio Maggiore. KURF2023 traced the incunabula from this library, providing detailed visualizations of the library's historical data and enhancing the platform's capability to track and analyze the historical trajectories of book collections.

OwnershipTracker (2024): The latest addition to the BookTracker platform, OwnershipTracker (Fig 2 f), was developed to visualize the complex networks of book ownership. This project was part of an MSc project by a student who previously contributed to KURF2023 and would like to explore this topic more. The resulting tool applied advanced visualization techniques to represent the complexities of social networks and ownership sequences, uncovering patterns and connections over time.

BookTracker continues to expand, with new projects like KURF2024 focusing on fresh aspects. This ongoing evolution underscores the dynamic and collaborative nature of the project, continually addressing new research challenges and leveraging the rich dataset provided by the MEI database.

4.2. Motivation for Development

Several motivations drove the continuous development of BookTracker. One important factor is the richness and complexity of MEI data. The MEI database offers a vast and detailed dataset; exploring this data from multiple perspectives can yield diverse and significant findings. This complexity necessitates specialized tools to fully leverage its potential, making comprehensive visualization and analysis essential.

MEI data users' varying needs and requirements have also driven the development of specialized visualization tools. For instance, historians might search the MEI database for data related to their specific research topics and need visualizations to validate their hypotheses. Data entry personnel use visual tools to identify incomplete entries in the database, aiding in future data completion and accuracy. Additionally, educators incorporate MEI data into their curricula, necessitating intuitive visualizations and interactive tools to communicate the information in classroom settings. By developing tools that address these specific needs, we enhance the usability and impact of the MEI data.

Certain historical research projects and specific research focuses have propelled the development of unique visualization tools within BookTracker. For example, in 2021, a comprehensive census of all surviving copies of Dante's Commedia's first edition, printed in 1481, was undertaken to commemorate its 540-year history and impact. This initiative directly led to the development of DanteSearchVis and DanteExploreVis. Similarly, studying the impact of book dispersal on religious orders and the effects of Napoleon's conquests on book migrations has driven the creation of special-

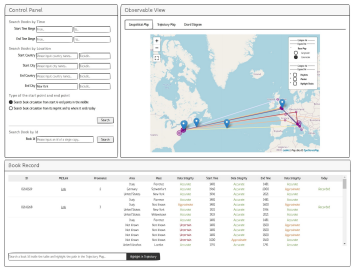
ized applications. These projects underscore the importance of developing targeted tools to uncover insights into specific historical phenomena.

These motivations highlight the necessity for a versatile and evolving platform like BookTracker to address the multifaceted nature of historical book trade research.

4.3. Components of BookTracker

Each component of BookTracker was developed to address specific domain problems. The components employ various visualization methods to provide insights into the historical book trade.

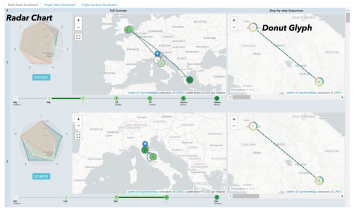
4.3.1. DanteSearchVis



The objective of DanteSearchVis is to search for and visually analyze book movement through various locations. This tool utilizes the MEI records entered for the census data of surviving copies of the first illustrated edition of Dante's Commedia, printed in 1481 and now scattered across 135 libraries worldwide.

DanteSearchVis employs various visualization techniques integrated into a cohesive interface consisting of three main components: a Control Panel, an Observable View, and a Record Table. The Control Panel facilitates query-based searching, allowing users to specify periods for the starting, ending, or any point in the middle of a book's life journey. The Observable View provides visual exploration through different visualizations, such as a geopolitical map and a chord diagram. The geopolitical map features interactive and animated paths depicting books' geographical movement. Users can interact with the map using various buttons for zooming, color scaling, and displaying polylines with arrows to indicate the direction of movement. This interactivity helps users grasp the flow and direction of book movements over time. Some of the project's findings have been summarized and published in [XDBAR22].

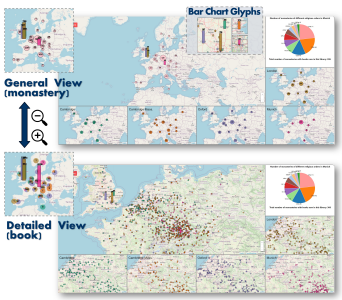
4.3.2. DanteExploreVis



DanteExploreVis aims to explore, explain, and present temporal and geographical data from book provenance records. As a sequential project following DanteSearchVis, it utilizes the same dataset but shifts the research focus towards extracting insights and exposing uncertainties within the data to domain experts.

DanteExploreVis employs several advanced visualization techniques, organized into three main panels: 1) the Query Panel enables initial data exploration and queries through heatmaps, which effectively show the frequency and distribution of book transfers, providing a global overview; 2) the Information Panel displays search results and highlights specific records, aiding users in pinpointing relevant information; and 3) the Storyboards represent the core visualization component, supporting data exploration and explanation through static and dynamic visualizations. Three distinct storyboards are designed to visualize individual copy records, group copy records, and movements through animations. Radar charts display data completeness, giving each record a quick visual gauge of information availability. Full journey maps visualize provenance and transfers on a geopolitical map, illustrating the geographical movement of books. Horizontal journey maps focus on the temporal progression of book movements, providing a timeline-based perspective. Additionally, donut charts are overlaid on the maps to illustrate the level of uncertainty associated with each provenance, allowing users to assess the reliability of the information. Some of the project's findings have been summarized and published in [XDBAR24].

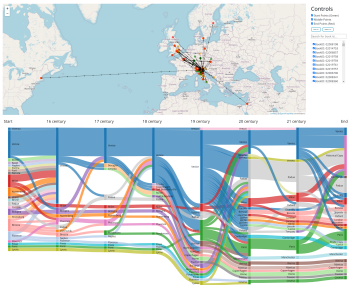
4.3.3. KURF2022



KURF2022 aims to investigate the dispersal patterns of books from religious orders throughout Europe during the sixteenth century and beyond. This project utilized a dataset of 2369 book records selected by domain experts from the MEI database, focusing on the impact of the secularization of religious houses on historical book dispersal.

KURF2022 uses map-based visualizations to display the dispersal of books, offering different levels of granularity to represent both monastery and book distributions. Users can zoom in and out to explore these levels, enabling detailed analysis at various scales. Additionally, small multiples [vdEvW13] are employed to synchronize and compare the holdings of five different libraries, allowing for comprehensive data comparison. This technique provides an intuitive way to understand the distribution of books across multiple institutions simultaneously. Bar chart glyphs [BKC*13] are integrated into the maps to represent the current libraries' holdings, distinguishing the five libraries and indicating the number of books each holds. These glyphs and synchronized mini-maps highlighting data for each library facilitate a detailed and comparative exploration of book origins and distributions.

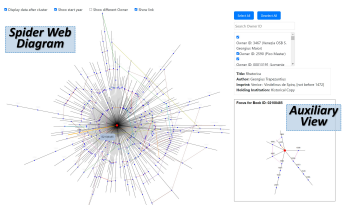
4.3.4. KURF2023



KURF2023 was developed to visualize the formation and subsequent dispersal of specific libraries due to historical events. This project focused on the San Giorgio Maggiore library in Venice, which included 181 incunabula now located in 35 libraries worldwide.

KURF2023 continues to employ interactive map visualizations to trace the movement of books from the San Giorgio Maggiore library. With a focus on the convergence and dispersal phenomena, a Sankey diagram is applied to illustrate the timeline and patterns of book circulation. The Sankey diagram allows users to click links to find corresponding book IDs, observe the timeline, and analyze circulation trends through link widths, reflecting the number of books in circulation between cities over time. The tool also includes search functionality, enabling users to find specific books using a search bar and view detailed circulation paths for selected books. This feature provides a quick understanding of book circulation by showing overall patterns without being distracted by intermediate circulations. By combining these techniques, KURF2023 offers a comprehensive and accessible way to explore the historical dispersal of books.

4.3.5. OwnershipTracker



OwnershipTracker was developed to explore and visualize the complex networks of historical book ownership and book transfer among various individuals and institutions. This tool leverages comprehensive provenance data from the MEI database, including ownership records and institutional holdings.

OwnershipTracker employs advanced network visualizations to represent the relationships between owners and holding institutions, and the books in their possession. The tool utilizes directed network graphs with interactive features to show the connections among various owners and institutions. Network graphs are particularly effective for illustrating complex relationships and connections, helping users understand patterns and sequences in book ownership. A spiderweb-like visualization, designed in collaboration with domain experts, presents the chronological order of provenance and inter-correlations between ownership records. Each radial axis represents a book's provenance extending outward, while the web-like connections between axes indicate relationships between different books' provenances and their owners. This detailed view of networks and ownership trajectories over time provides significant insights into the historical context of book distribution and possession.

4.4. Design and Evaluation Methods

The design and evaluation of the BookTracker platform were guided by a collaborative approach with domain experts to ensure that the tools effectively met their needs. The development process followed a typical visualization collaborative design study, incorporating principles from the Nine-Stage Framework [SMM12], which was adapted to better fit the iterative nature of our projects.

While collaborating with historians, we identified key aspects such as the significance of clear communication and understanding of academic terminology in cross-disciplinary collaborations, the strong divergent thinking abilities of humanities researchers, and

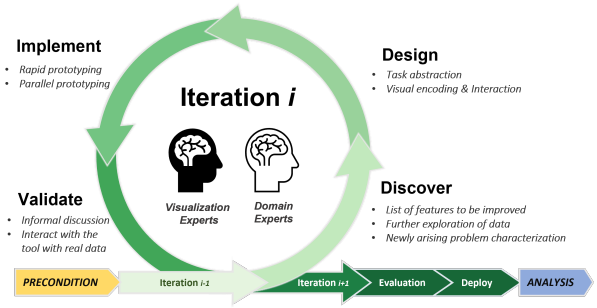


Figure 3: *This figure illustrates the iterative design methodology we follow, adapted from the Nine-Stage Framework [SMM12].*

the evolving perceptions of data and requirements as domain experts engaged with visualization tools. These insights informed our approach to design and evaluation, ensuring a more user-centered process.

We adapted the Nine-Stage Framework, which is primarily linear with backward reflective loops, by tailoring its core phases (*discover*, *design*, *implement*, *deploy*) into iterative cycles. The first three stages were retained as part of prototype development cycles, with weekly meetings serving as iteration nodes. A validation stage was added within each iteration, facilitating controllable evaluation practices with domain experts and ensuring regular communication throughout the design process. The Agile model of software development [ELS05] also inspired this adaptation, emphasizing user engagement and increasing collaboration between both parties (historians and visualization researchers) during the core development stages.

The framework variant is depicted in Fig. 3. The *precondition* and *analysis* phases were adopted from the original framework, while the core phase was restructured into iterative cycles with modified objectives. This approach allowed for the rigorous execution of evaluation and reflection activities through iteration, maintaining frequent and consistent interactions with domain experts.

The evaluation of the BookTracker components varied depending on the scope and duration of each project. For DanteSearchVis and DanteExploreVis, we conducted systematic evaluations, usability testing, and expert feedback to ensure the tools met the high standards required for academic research. For other projects, the testing and evaluation were constrained by the project timelines. We relied on domain expert feedback and an incremental development approach. Regular feedback sessions with historians and other domain experts helped refine the tools, ensuring they effectively addressed the evolving research needs.

4.5. Integration and Cohesion

The BookTracker platform integrates these tools into a cohesive system supporting the comprehensive analysis of MEI records. The platform's evolution reflects a gradual expansion from simple geographical-oriented visualizations to more complex analyses of different historical contexts. Each tool builds on the capabilities

of its predecessors, adding new dimensions and perspectives to the analysis of historical book trade data.

The platform's development has been characterized by a strong emphasis on collaboration with domain experts, ensuring that each tool meets the specific needs of researchers. This long-term collaborative effort has been instrumental in refining and enhancing the platform, making BookTracker an indispensable resource for historical book trade research. By exploring the MEI data from multiple angles, researchers can uncover diverse and significant findings, addressing MEI data users' complex and varied requirements. Specific historical research projects have further driven the development of targeted tools, highlighting the importance of specialized applications in uncovering insights into historical phenomena.

5. Reflections and Discussions

5.1. The Importance of Long-Term Collaboration

The role of visualization technology is critical in the workflow of humanities scholars, serving not to replace human effort or alter established methodologies but to complement and enhance the exploration and presentation of complex datasets. Effective visual analytics tools should focus on helping scholars manage large volumes of information, detecting significant patterns in extensive datasets, and presenting their findings compellingly for research and educational purposes. Moreover, these tools are expected to provide unbiased visual evidence to support or challenge scholars' intuitive hypotheses, significantly aiding their scholarly inquiries.

In domain problem-focused visualization design studies, it is common for domain experts new to visualization technologies to need time to understand and fully integrate these tools into their workflows. The value of human-tool collaboration emerges gradually, necessitating persistent cooperation between domain experts and visualization researchers. This relationship thrives on iterative development and feedback, enhancing tool utility and integration into established practices. However, the literature, including many design study papers, frequently points out the challenges in bridging the comprehension gap between these groups [BSS20, BEAC*18]. Effective bridging strategies involve immersing both parties in each other's fields and adopting a human-centered design approach. While this reduces barriers to understanding, it requires time and sustained engagement.

Unfortunately, too many projects conclude prematurely–often after the publication of a single paper or the graduation of a PhD student–leading to a squandered opportunity to build on the collaborative knowledge gained. It is crucial to transition these initial collaborations into long-term projects to prevent such loss and foster ongoing innovation. This approach, however, must be balanced with practical considerations such as funding and resource allocation. By sharing the development process of the BookTracker platform, we hope to demonstrate our strategies for maintaining sustainable research collaborations, providing a case study for others facing similar situations. We also aim to open the discussion for further exploration.

Through long-term collaboration, visualization researchers and humanities scholars can continuously refine their mutual understanding, enhancing the tools they develop and extend. Such persistent efforts lead to more impactful research outcomes and ensure that visualization tools are deeply integrated into and evolve with the humanities workflow, meeting the ever-changing needs of researchers and educators.

5.2. Strategies for Sustaining Long-Term Collaboration

While the importance of long-term interdisciplinary collaboration is well recognized, it poses significant challenges. Sustained engagement, personnel shifts, and financial factors all represent potential constraints. The BookTracker platform started as a PhD project, with DanteSearchVis being the first collaborative visualization tool. From our experience, we summarized several factors that contributed to the project's evolution from a single initiative to an expanding platform with multiple sub-projects.

Domain Needs as a Driving Force. The most critical factor driving the project's development is the domain needs. In our case, the field of historical book research, traditionally less exposed to visualization techniques, presents abundant opportunities for employing visualization research to support scholarly work. Under the auspices of the 15cBOOKTRADE Project, provenance data are continuously aggregated from multiple libraries and book collections. Domain experts utilize this data and material evidence from the books to reconstruct the history and "life" of books. As the data corpus grows, so does the demand for technological tools (visualizations and data analytics). From the perspective of domain experts, the expanding datasets provide a growing mass of material evidence to substantiate their research hypotheses. The need for visualization support varies significantly based on their research focus, though the data structure remains similar. For instance, understanding the distribution of a specific writer's books and the phenomena of these books' circulation or the imprint of books from a specific period influenced by a particular historical event. The overall structure of the data remains consistent (provenance objects with geographical and temporal features), yet the subsets of data differ and are imbued with specific domain contexts, thus independently forming research questions. These can be studied and addressed using different visualization and visual analytic techniques. This domain specificity creates a robust demand for visualization, serving as a foundational support for our sustained collaboration.

Availability, Completeness, and Data Accessibility. A crucial supporting factor in our project has been the MEI database. This database, constructed and maintained by domain experts, is a comprehensive collection of provenance data. Despite the inherent uncertainties within the data, these uncertainties are manageable and do not significantly affect the effectiveness of visualization outputs. Moreover, the MEI database includes a robust API system for data retrieval and access, allowing visualization developers to familiarize themselves with the data and begin their work quickly. This accelerates the initial stages of each project, providing a solid foundation for development. Overall, the richness of the MEI data supports multiple sub-projects and allows for easy data acquisition, enabling computer science and visualization researchers to customize their queries to obtain the required data easily.

Continuous Supply of Developers (Visualization Researchers).

Another significant factor is the continuous influx of developers and visualization researchers. Our project's core team of visualization experts is not static; it changes with different project phases. The lead developers often vary, encompassing PhD students, master's students, and undergraduate students. Their involvement in the project stems from various motivations–some are participating as part of their coursework, others are applying classroom-learned software development and visualization knowledge in a practical setting, and a genuine interest in interdisciplinary projects and historical books drives some. This dynamic and fluid development team has sustained the project's vitality and enhanced its impact.

Enthusiasm for Collaboration (Domain Experts). Another valuable aspect of our project has been the sustained engagement of the lead domain expert. Over nearly four years, she has actively participated in co-designing, discussing, and testing activities. She has gradually familiarized herself with our interdisciplinary workflow and adapted to our methodologies and ideas, particularly in proposing research questions and related data scopes. The domain expert proposes questions of interest along with the scope of data retrieval pertinent to the historical research questions and explains the significance of these research questions to visualization researchers from an accessible perspective, highlighting areas that significantly benefit from visualization support. Moreover, she has mastered previous participants' common misunderstandings and emphasizes these points during the initial discussion phases. This experience has significantly accelerated the establishment of mutual knowledge transfer in subsequent projects. While a dynamic and high-turnover team of developers is crucial, the continued support from a dedicated domain expert is equally important for fostering long-term interdisciplinary collaboration.

Enhancing Participation through Design Study. During each project, we believe that adherence to a specific workflow (the design study model introduced in Section 4.4) greatly aided the progression and long-term sustainability. We adapted the Nine-Stage Framework and incorporated an Agile development structure, highlighting the critical role of iteration in the design and development phases. By conducting regular face-to-face interactions (either weekly or bi-weekly), we ensured that domain experts were actively engaged with the evolving prototypes, allowing them to test and provide feedback extensively. This approach significantly enhanced the domain experts' participation and sense of reward. Simultaneously, for visualization researchers and tool developers, these explicit iterative feedback sessions—through regular discussion meetings—improved their opportunities for interaction with domain experts and maintained a consistent development pace. Such a project execution strategy enhances the sense of engagement and integration for all parties involved and ensures the project progresses at a steady rhythm.

5.3. Benefits for Both Parties

Jänicke argues that the value of visualization developed in a digital humanities project is not always equally high for both research fields. He gives an example of this imbalance: *"When an approach is too complex – which counts as a strong argument for a publication in a visualization realm – it might become invaluable for humanities scholars due to comprehension problems. On the other hand, if a clean, easily understandable visualization is valuable for a humanities scholar, the missing novelty most likely impedes a computer science publication"* [Jän16]. We recognize this viewpoint and have encountered similar situations in the development of BookTracker. However, achieving mutual benefits is crucial for a long-term cross-disciplinary collaboration to be sustainable. Beyond the mismatch in research publication priorities, can visualization researchers gain from such activities? From our experience, we believe the answer is yes.

Working closely with humanities scholars presents unique and challenging problems for visualization researchers. This collaboration pushes the boundaries of existing visualization techniques and leads to the development of innovative solutions or adapting existing approaches to suit new application fields. The interdisciplinary nature of the work provides valuable experience, enriching the researchers' understanding of how visualization can be applied in various domains. Additionally, it fosters a deeper appreciation for the complexities and nuances of humanities data, creating more versatile and robust visualization tools. These projects offer novice visualization researchers an excellent opportunity to enhance their skills in creating visualization applications. These collaborations often open up new research avenues and opportunities for funding, expanding the scope and impact of their work.

On the other hand, humanities scholars benefit significantly from long-term collaborations by gaining powerful tools to analyze and present their data more effectively. Visualization techniques help uncover patterns and insights that are not immediately apparent through traditional methods. These tools also enhance their ability to communicate research findings to a broader audience, including students and the general public. The visualizations make complex data more accessible and engaging, facilitating better understanding and retention of information. Moreover, the iterative development process ensures that the tools are continuously refined to meet the evolving needs of humanities research, ultimately leading to more accurate and insightful analyses.

In summary, while the value balance between the two fields may not always be equal, the long-term benefits of sustained collaboration are significant. Visualization researchers gain new challenges and insights, driving innovation and expanding their expertise. Humanities scholars receive advanced tools that enhance their research capabilities and communication. This mutual benefit is essential for maintaining and growing cross-disciplinary projects like BookTracker.

6. Conclusion

The BookTracker platform exemplifies the benefits of long-term interdisciplinary collaboration between visualization researchers and humanities scholars. Utilizing the comprehensive data from the MEI database, BookTracker's tools have significantly enhanced the analysis of historical book trade, from tracking geographical movements to analyzing ownership networks. Domain experts already use the tool for presenting and teaching, showcasing its practical applicability. In summary, BookTracker demonstrates how interdisciplinary efforts can create powerful visualization tools, driving innovation and expanding research capabilities in both visualization and humanities.

Acknowledgements

This project is partly funded by the King's-China Scholarship Council PhD Scholarship programme (K-CSC). We would like to express our sincere gratitude to Meilai Ji, Xiangyi Liu, and Dengyi Yan, undergraduate students supported by the King's Undergraduate Research Fellowships (KURF), for their dedicated efforts in developing BookTracker's components. Additionally, we extend our heartfelt thanks to all the domain experts who have significantly contributed to the evaluation of BookTracker.

References

[AKV*14] ALEXANDER E., KOHLMANN J., VALENZA R., WITMORE M., GLEICHER M.: Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 173–182. doi:10.1109/VAST.2014.7042493. 2

[ARRO*17] ABDUL-RAHMAN A., ROE G., OLSEN M., GLADSTONE C., WHALING R., CRONK N., MORRISSEY R., CHEN M.: Constructive Visual Analytics for Text Similarity Detection. *Computer Graphics Forum* (2017). doi:10.1111/cgf.12798. 2

[BEAC*18] BRADLEY A. J., EL-ASSADY M., COLES K., ALEXANDER E., CHEN M., COLLINS C., JANICKE S., WRISLEY D. J.: Visualization and the digital humanities: Moving toward stronger collaborations. *IEEE Comput. Graph. Appl. 38*, 6 (2018), 26–38. doi:10.1109/MCG.2018.2878900. 2, 7

[BKC*13] BORGO R., KEHRER J., CHUNG D. H. S., MAGUIRE E., LARAMEE R. S., HAUSER H., WARD M., CHEN M.: Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics 2013 - State of the Art Reports* (2013), Sbert M., Szirmay-Kalos L., (Eds.), The Eurographics Association. doi:10.2312/conf/EG2013/stars/039-063. 5

[BSS20] BENITO-SANTOS A., SÁNCHEZ R. T.: A data-driven introduction to authors, readings, and techniques in visualization for the digital humanities. *IEEE Computer Graphics and Applications 40*, 3 (2020), 45–57. doi:10.1109/MCG.2020.2973945. 2, 7

[CARM17] COLE N., ABDUL-RAHMAN A., MALLON G.: Quill: A framework for constructing negotiated texts - with a case study on the us constitutional convention of 1787. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (2017), pp. 1–10. doi:10.1109/JCDL.2017.7991562. 2

[CER15] CERL C.: Material Evidence in Incunabula, 2015. URL: https://data.cerl.org/mei/_search. 1, 2

[CMW*17] CHANDLER T., MCKEE B., WILSON E., YEATES M., POLKINGHORNE M.: A new model of angkor wat: Simulated reconstruction as a methodology for analysis and public engagement. *Australian and New Zealand Journal of Art 17*, 2 (2017), 182–194. doi:10.1080/14434318.2017.1450063. 2

[Don13] DONDI C.: "15cBooktrade": An evidence-based assessment and visualization of the distribution, sale and reception of printed books in the renaissance. *Gazette du livre médiéval 60*, 1 (2013), 83–101. 2, 3

[Don20] DONDI C.: *Printing R-Evolution and Society 1450-1500*. Edizioni Ca' Foscari, feb 2020. doi:10.30687/978-88-6969-332-8. 2, 3

[Don24] DONDI C.: From the 15cBOOKTRADE Project website, (accessed in 2024). URL: https://15cbooktrade.ox.ac.uk/. 2, 3

[Dru11] DRUCKER J.: *Humanities Approaches to Graphical Display*. Wiley Online Library, 2011, ch. 17. doi:10.1002/9781118680605.ch17. 2

[EAHJG*17] EL-ASSADY M., HAUTLI-JANISZ A., GOLD V., BUTT M., HOLZINGER K., KEIM D.: Interactive visual analysis of transcribed multi-party discourse. In *Proceedings of ACL 2017, System Demonstrations* (jul 2017), Association for Computational Linguistics, pp. 49–54. URL: https://aclanthology.org/P17-4009. 2

[ELS05] ERICKSON J., LYYTINEN K., SIAU K.: Agile modeling, agile software development, and extreme programming: the state of research. *Journal of Database Management (JDM) 16*, 4 (2005), 88–100. 6

[Jän16] JÄNICKE S.: Valuable research for visualization and digital humanities : A balancing act. In *Workshop on Visualization for the Digital Humanities, IEEE VIS 2016, Baltimore, Maryland, USA* (2016). URL: https://api.semanticscholar.org/CorpusID:40584357. 8

[JW17] JÄNICKE S., WRISLEY D. J.: Interactive visual alignment of medieval text versions. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2017), pp. 127–138. doi:10.1109/VAST.2017.8585505. 2

[KJW*14] KOCH S., JOHN M., WÖRNER M., MÜLLER A., ERTL T.: Varifocalreader — in-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1723–1732. doi:10.1109/TVCG.2014.2346677. 2

[MLCM16] MCCURDY N., LEIN J., COLES K., MEYER M.: Poemage: Visualizing the sonic topology of a poem. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 439–448. doi:10.1109/TVCG.2015.2467811. 2

[PHS24] PRICKMAN G., HOLLAND A., SHEPARD R.: The Atlas of Early Printing, (accessed in 2024). URL: https://atlas.lib.uiowa.edu/. 3

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2431–2440. doi:10.1109/TVCG.2012.213. 6

[THC12] THUDT A., HINRICHS U., CARPENDALE S.: The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, Association for Computing Machinery, p. 1461–1470. URL: https://doi.org/10.1145/2207676.2208607, doi:10.1145/2207676.2208607. 2

[vdEvW13] VAN DEN ELZEN S., VAN WIJK J. J.: Small multiples, large singles: A new approach for visual data exploration. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 191–200. 5

[Wri17] WRISLEY D.: Locating medieval french, or why we collect and visualize the geographic information of texts. *Speculum 92*, S1 (2017), 145–169. doi:10.1086/691536. 2

[Wri20] WRISLEY D.: Digital spatial practices and linguistic landscaping in beirut. *ILCEA*, 39 (March 2020). doi:10.4000/ilcea.9198. 2

[XDBAR22] XING Y., DONDI C., BORGO R., ABDUL-RAHMAN A.: A Design Study of Visualizing Historical Book Movement. In *EuroVis 2022 - Short Papers* (2022), Agus M., Aigner W., Hoellt T., (Eds.), The Eurographics Association. doi:10.2312/evs.20221100. 5

[XDBAR24] XING Y., DONDI C., BORGO R., ABDUL-RAHMAN A.: Visualizing historical book trade data: An iterative design study with close collaboration with domain experts. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (2024), 540–550. doi:10.1109/TVCG.2023.3326923. 5