An Empirical Study on Personal Information Management Practices Across Scholarly Workflows

Qi Xu
Web & Information Systems Engineering Lab
Vrije Universiteit Brussel, Pleinlaan 2
Brussels, Belgium
qi.xu@vub.be

Beat Signer
Web & Information Systems Engineering Lab
Vrije Universiteit Brussel, Pleinlaan 2
Brussels, Belgium
bsigner@vub.be

Abstract

Contemporary scholars navigate complex workflows that involve managing vast amounts of information, from literature and experimental data to manuscript writing. While prior research in Scholarly Personal Information Management (Scholarly PIM) shed light on scholarly information management practices, it has yet to fully incorporate the concept of scholarly workflows into its analytical framework. Our study addresses this gap by combining a user survey with guided tour interviews to investigate how scholarly PIM practices work across scholarly activities. By introducing the concept of scholarly workflows into scholarly PIM studies, our preliminary findings provide an empirically grounded view of contemporary scholarly PIM practices and challenges. Our results show that framing scholarly PIM within the full spectrum of scholarly work advances the understanding of scholarly PIM from a workflow-oriented perspective and might inspire the design of future systems and user interfaces that better support scholars' evolving information management needs.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Information systems**;

Keywords

Personal information management, PIM, scholarly workflow

ACM Reference Format:
Qi Xu and Beat Signer. 2026. An Empirical Study on Personal Information Management Practices Across Scholarly Workflows. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3772363.3798664

1 Introduction

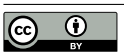
Throughout their careers, scholars continuously gather and build complex personal collections of research material such as printed or digital archival documents [9]. As these collections expand, managing them efficiently becomes increasingly difficult, leading scholars to gradually lose control over their research-related material. Current research in this domain primarily stems from a subfield of Personal Information Management (PIM) known as Scholarly Personal Information Management (Scholarly PIM) [6, 8, 19, 26], contextualising PIM theory within an academic setting. It provides an initial understanding of the fundamental concepts of scholarly PIM, scholars' information management strategies and tools, and the limitations of their current practices.

Scholarly workflows refer to the series of interconnected scholarly activities that span the research lifecycle, from finding, storing and analysing to writing, sharing and publishing [25]. However, current research on scholarly PIM has yet to fully incorporate this concept into its analytical framework. Our research objective is to address this gap and investigate how scholarly PIM practices work across scholarly workflows. To address this research objective, we employed a mixed-methods approach combining a user survey with a guided tour study. We asked participants to walk us through their scholarly workflows and gathered data on their scholarly PIM practices at each stage of the workflow. Our findings provide an empirical, though preliminary, overview of current PIM practices across entire scholarly workflows, highlighting strategies, tools and challenges in managing diverse types of scholarly information.

2 Related Work

Personal Information Management (PIM) is broadly defined as the practice and study of activities that individuals undertake to acquire, store, organise, maintain, retrieve, use and distribute information needed to meet life's many goals [14]. Unlike conventional information management, where information is organised by its intrinsic properties, PIM is typically structured around users' purposes and tasks [10]. The most widely accepted PIM model by Jones [14] conceptualises PIM as an ongoing effort to map information to personal needs through three primary Personal Information Management Activities (PIM Activities): *"finding/re-finding"*, *"keeping"* and *"meta-level"* activities. Meta-level activities include *"organising"*, *"maintaining"*, *"measuring and evaluating"*, *"managing privacy and the flow of information"* and *"sensemaking"*. Personal Information Management Practices (PIM Practices) refer to techniques, methods or strategies of PIM that were initially selected by the person to accomplish some purpose and are now a part of the person's PIM repertoire [15]. PIM practices can be seen as the concrete, situated manifestations of the more abstract concept of PIM sactivities.

A *scholarly workflow* is a series of purposeful processes for scholars to conduct research from scratch [25]. These processes are also known as scholarly activities, and different scholarly workflow models have different definitions and classifications, such as finding, storing and analysing information, as well as writing, sharing



CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/2026/04
https://doi.org/10.1145/3772363.3798664

and publishing research outputs. Many scholars have examined various aspects of scholarly workflows to develop models that describe their structure and explain the interactions among their components. There are five widely accepted scholarly workflow models proposed by different researchers, comprising Palmer et al.'s [20] model, Antonijević's [2] model, Bosman and Kramer's [5] model, Cooper and Reige's [9] and Ince et al.'s [13] model. Although all these studies focus on scholarly workflows, their emphasis and analytical perspectives differ. Some models, such as the ones by Antonijević or Ince et al., place greater emphasis on scholarly activities related to the disciplines of humanities and social sciences. Others, such as Bosman and Kramer's model, place greater emphasis on the role of tools and technologies in scholarly workflows.

The concept of *Scholarly Personal Information Management (Scholarly PIM)* builds on the general PIM framework and focuses on managing scholarly information. Although the source of scholarly information is often non-personal (public or shared), scholars engage with it for personal purposes, such as their individual scholarly career [10], and manage it on a personal level [6]. Even when engaged in collaborative research, scholars typically manage their own material independently [1], which can be considered a form of PIM. Bussert et al. [6] described scholarly PIM as a distinct form of PIM tailored to the needs of researchers. According to Bussert et al., scholarly PIM refers to the management of information directly related to scholarly tasks undertaken by individuals such as researchers and graduate students.

A growing body of research has examined scholarly PIM practices [3, 6, 8, 9, 11, 19, 20, 24, 26]. Existing studies generally follow a similar approach: they begin by identifying common strategies associated with the six PIM activities of *finding*, *re-finding*, *keeping*, *organising*, *maintaining* and *sensemaking* in scholarly contexts. For instance, note-taking is often considered a typical strategy for the PIM activity of sensemaking in academic work. From these strategies, researchers identify the PIM tools used, such as Citation Management Tools (CMTs). Further, they explore the limitations and challenges of current practices, including difficulties in managing both physical and digital material.

However, existing research on scholarly PIM practices faces a major limitation: the lack of investigation into how PIM activities work across entire scholarly workflows. Existing studies rarely examine scholars' PIM practices in a stage-by-stage manner throughout their workflows. In other words, little is known about how scholars manage scholarly information within and across individual scholarly activities, as well as how these practices vary or evolve across stages. This gap raises critical questions that merit rigorous, systematic investigation: What are the current practices that scholars use to manage scholarly information across scholarly activities? Further, what challenges do they encounter in the process of conducting PIM activities at each stage of their workflows?

3 Method

We adopted a mixed-methods research approach consisting of a questionnaire-based user survey and a follow-up guided tour study. The questionnaire-based *user survey* was designed and distributed to participants—all of whom were scholars, including undergraduate and master's students, PhD candidates, postdoctoral researchers, research staff, academic faculty and industry researchers—using the Qualtrics platform. The survey consisted of four sections: (1) demographic and background information, including gender, academic status and research discipline (Appendix A.1); (2) participants were introduced to terminology covering 40 scholarly activities derived from five established workflow models [2, 5, 9, 13, 20] and asked to rate how accurately each term reflected their own scholarly workflows (Appendix A.2); (3) from the set of 40 scholarly activities introduced in the previous section, participants were asked to select those activities that best reflected their current practices in order to construct a personalised scholarly workflow model based on existing models (Appendix A.3); (4) participants were asked to describe their PIM practices step by step based on their own scholarly workflow models created earlier (Appendix A.4). Participants who volunteered for a follow-up study took part in a guided tour, conducted online or in person, with audio and screen recordings captured via Microsoft Teams. During the study, participants were asked to share their screens to demonstrate their actual working environments and walk through their entire scholarly workflow. For each scholarly activity, they were instructed to show how they use existing scholarly PIM tools to perform relevant PIM activities and to explain any challenges they encounter in the process. After each phase of the demonstration, we asked some follow-up questions and made targeted requests for additional demonstrations to gain a deeper insight into their scholarly PIM practices. Each interview lasted approximately 20 to 40 minutes.

4 Preliminary Findings

Our findings are based on a first round of data collection and NVivo-based thematic analysis of survey responses, video recordings and transcribed audio data from 28 survey participants and 10 guided tour participants. Participants represented diverse disciplinary backgrounds and scholarly statuses, with balanced gender representation (14 male, 14 female) and an age range spanning 18–64 years (18–24: 5; 25–34: 12; 35–44: 4; 45–54: 5; 55–64: 2). In terms of disciplinary background, 18 participants came from Engineering and Technology, 6 from Social Sciences, 1 from Natural Sciences, 1 from Humanities and 2 "other". Regarding their scholarly status, 1 was a Postdoc, 12 were PhD candidates, 2 were Bachelor's students, 7 were Master's students, 3 were faculty members, 2 were industry researchers and 1 "other". Further, in the guided tour study, 9 participants were PhD candidates and 1 was a Postdoc.

4.1 Common Types of Scholarly Activities

Through analysis of coded survey and interview data, we identified the three common scholarly activity types of *collecting*, *curating* and *creating* shown in Table 1. Collecting scholarly activities focus on gathering raw primary material (e.g. experimental data, survey data and interview data) and secondary material (e.g. papers, books, news and social media content) by searching external sources, extracting relevant information and storing it locally or remotely for future retrieval. Curating scholarly activities centre on interpreting, analysing and processing raw material obtained during collecting activities to derive meaningful insights. In creating scholarly activities, scholars transform material collected during collecting and

Type	Collecting	Curating	Creating
Scholarly Activities	Searching [20] Collecting [20] Monitoring [20] Collect [2] Find [2] Discovery [5] Collecting data, evidence or insights [9] Information Literacy [13]	Reading [20] Analyse [2] Annotate [2] Analysis [5] Analysing data, evidence or insights [9] Knowledge Management [13]	Writing [20] Write [2] Cite [2] Writing [5] Publication [5] Outreach [5] Writing, presenting, sharing, engaging, seeking feedback [9] Grant writing and fund raising [9] Interpreting findings and testing hypotheses [9]
	Formulating research goals [9] Selecting data gathering and experimental methods [9]		
	Notetaking [20], Translating [20], Data Practices [20], Collaborating [20], Organise [2], Reflect [2], Archive [2], Communicate [2], Share [2], Assessment [5], Conducting a literature review [9], Assessing findings and conducting quality control [9], Collaborating or networking [9], Information Management [13], Scholarly Communication [13]		

Table 1: Scholarly activities grouped by the three identified types of *collecting, curating* and *creating*

analytical intermediate information from curating into mature research output. These three types of scholarly activities constitute a high-level categorisation that encompasses the major phases of a scholarly workflow. Notably, some scholarly activities span multiple types, such as 'Translating' [20], which occurs across all three. These activities correspond to what Palmer et al. [20] describe as *"cross-cutting"* scholarly activities, which are commonly co-present alongside many other scholarly activities.

4.2 Current Scholarly PIM Practices and Their Limitations Across Scholarly Workflows

In Section 4.1, we identified three common types of scholarly activities covering the typical scholarly workflow, and subsequent analysis of the coded data showed that PIM practices within the same activity type often exhibit strong similarities. Thus, we categorised scholarly PIM practices into three groups: PIM practices related to collecting, curating and creating scholarly activities. The findings presented below enable an initial, holistic examination of scholarly PIM practices across entire scholarly workflows.

4.2.1 PIM Practices Related to Collecting Scholarly Activities. For the *keeping*, primary material is typically stored as files in research-specific formats (e.g. CSV or mp4) on local or remote file systems. Secondary material, such as academic papers (usually PDFs), is most commonly saved in file managers or in CMTs such as Zotero (Appendix B.1). The practices of *organising* and *maintaining* are closely intertwined with *keeping*, as storage locations largely determine how information is organised and maintained, and these processes across different storage locations remain independent. For example, although PDF articles are organised and maintained using folder hierarchies in both CMTs and file managers, these hierarchies are not synchronised. As a result, scholars often replicate similar folder structures and manage them separately. *Re-finding* commonly relies on search, tagging or navigating structures such as folder hierarchies. As can be observed, in collecting scholarly activities, information is typically managed at the level of files or documents when engaging in PIM activities such as *keeping*, *organising*, *maintaining* and *re-finding*. With respect to *sensemaking*, scholars frequently reported that file managers and CMTs only provide limited support for making sense of collected files. A key reason is that scholars rarely engage in sensemaking at the level of entire files or documents; instead, they tend to interact with and make sense of more granular units of scholarly information.

4.2.2 PIM Practices Related to Curating Scholarly Activities. Based on our observation, scholars extract useful excerpts from source material, and the most commonly reported *keeping* practices for these excerpts fall into three strategies: (1) Annotations: excerpts are retained as highlights within original files (e.g. PDFs, text documents, images); (2) Notes: excerpts are kept in standalone notes through copy/paste, screenshots, links or paraphrasing; (3) Codes: excerpts are kept in specialised qualitative coding tools.

The *annotations strategy* is widely adopted. PDFs are the most frequently annotated material by scholars, with participants typically relying on PDF viewers embedded in CMTs or standalone ones (Appendix B.2). Because annotations are tightly bound to their source documents—being embedded within and structurally dependent on the original file formats—they offer limited flexibility for *organising* and *maintaining* annotations as independent information units; *re-finding* annotations is most commonly accomplished through direct search within the document. This is also why annotations inherently retain the excerpt's surrounding document-level context. However, this contextual preservation is confined to the annotated file itself. Scholars often rely on external material as background information during reading and analysis, and contextual information beyond the annotated file is therefore difficult to retain, as existing annotation systems typically provide only plaintext comment fields for recording external contextual information.

The *notes strategy* is the most commonly adopted, with participants using both general-purpose authoring tools, such as Microsoft Word, and dedicated note-taking tools such as Zotero's note module and Obsidian (Appendix B.3). In general-purpose tools like Word, excerpts are often stored via copy/paste, screenshots or paraphrasing. In contrast, in specialised tools such as Obsidian, excerpts are more commonly stored via links to parts of the source material. For *organising*, these tools typically provide file manager-like structures, including folders and tags. *Re-finding* similarly relies on these organisational mechanisms or on keyword search. Unlike annotations, excerpts in notes are decoupled from the original files. A major limitation of this approach is that contextual information, such as the original content, background information, and provenance, is often lost during note-taking. When scholars rely on copy/paste or screenshots, the original content is retained, but background and provenance information is often lost. For paraphrasing, all contextual information is missing because paraphrasing removes explicit informational anchors to the source. When links are used, contextual information can be preserved to some extent. However,

current linking mechanisms are not easy to use and typically support only PDFs, with limited or no support for other media types. As a result, excerpts stored in notes often remain disconnected from their original source material, creating difficulties for understanding coherence and tracking provenance as research progresses.

The *codes strategy* was mentioned the least often and NVivo was the only tool reported by participants within this strategy. Coding tools support *organising* and *maintaining* by grouping semantically related excerpts under shared codes. *Re-finding* can be achieved through search or by navigating code labels. Such tools preserve contextual information, but they are often perceived as heavyweight and cumbersome, leading to a less satisfactory user experience for everyday scholarly work.

All three strategies support *sensemaking* to some degree. During *sensemaking*, scholars often draw on a variety of external representational formats. The most common form is linear text, but some participants also employ structures such as mind maps, whiteboards, node-link diagrams and tables/spreadsheets (Appendix B.4). Scholars reported that external representations support the understanding of large bodies of material by enabling efficient identification of salient information and maintaining an overview.

4.2.3 PIM Practices Related to Creating Scholarly Activities. Personal information management practices related to *creating scholarly activities* rely on different infrastructures depending on scholars' authoring environments. When using local editors (e.g. Word), scholars typically depend on file managers for information management, whereas cloud-based editors (e.g. Overleaf) provide built-in document management functionality. Participants did not report major dissatisfaction with these PIM activities. However, during creating, re-finding the necessary reference metadata for citation was often cumbersome. Scholars were unable to directly access the corresponding source metadata from excerpts stored in notes; instead, they had to use the source title recorded in the notes to search for the reference again in a CMT in order to retrieve the required metadata, making citation inefficient and labour-intensive.

5 Discussion and Future Work

Based on a first round of data collection and analysis, we found that incorporating scholarly workflows yields a richer understanding of scholarly PIM practices and identified several noteworthy phenomena, as discussed below. In Section 4, we introduced and utilised the concept of "excerpts", which we now explore in greater detail. Our findings suggest that when scholars engage in deeper processing of scholarly information (curating scholarly activities), they tend to work with excerpts as their primary unit of interaction rather than with entire documents. This observation has also been raised by other researchers. Qian et al. [22] argue that a major limitation of current CMTs is that they are designed around the principle of treating research papers as the fundamental unit of interaction. While papers, as a fundamental unit, reflect the formal structure of scientific publishing, this may not align with how researchers actually engage with literature during reading, annotating and writing. In practice, scholars often need to work at a more granular level than on entire documents. As Al-Omar and Cox [1] note, *"the information in the document is more important than the document itself"*. In scholarly activities, scholars primarily interact with the information embedded in documents. For example, a common scholarly practice is to highlight and annotate excerpts of a research paper to extract and preserve key ideas by using tools such as a PDF highlighter [7]. Similarly, researchers sometimes use qualitative data analysis tools like NVivo to *code* excerpts from interview transcripts or observational data [17]. Our findings also show that a common practice among scholars is to store excerpts in notes for the purposes of analysis and understanding. These practices suggest that scholars may be engaging with more granular units of scholarly information, rather than treating the document or file as the unit of interaction.

Our findings indicate that excerpts are commonly managed through annotations, notes, and codes as information management structures, rather than through more commonly used structures such as folders or tags. While folders and tags, as conceptualised by Jones [14], serve effectively as "grouping items" for documents, they are ill-suited for managing excerpt-level scholarly information. In contrast, annotations, notes and codes emerge as the primary grouping item for excerpts, as they are better aligned with the PIM activity of sensemaking than folders or tags. They provide external representations such as schemas, diagrams, and tables that support scholars in organising, relating, and interpreting information [21, 23]. However, current excerpt-level PIM strategies provide limited support for efficiently preserving contextual information associated with excerpts, such as original content, background information and provenance. In PIM research, the subjective context principle suggests that information should be retrieved by the user in the same context in which it was previously used, in order to bridge the time gap between retrieval and prior use [4]. For scholars, such contextual continuity is essential for later understanding and reuse of information, especially in scholarly tasks like synthesis [18]. Although scholars can currently technically reconnect excerpts to source documents through creating links, these approaches are often perceived as cumbersome and burdensome in practice [12]. This underscores the need for efficient, seamless context-aware support in excerpt-level scholarly PIM tools. Overall, excerpt-level information management remains underexplored in scholarly PIM and calls for further investigation.

Our findings suggest that several PIM challenges observed across different stages of the scholarly workflow share a common underlying cause: information fragmentation. Unsynchronised folder structures during collecting, difficulties in maintaining contextual information during curating, and the cumbersome re-finding reference metadata during creating, largely arise because scholarly information is managed independently across different PIM tools, which have limited ability to access or interoperate with one another. Our analysis indicates that scholars rely on distinct PIM tools, each of which employs its own information management structures (grouping items) to organise content. These structures are often incompatible, leading each tool to manage information in isolation. As a result, scholars face substantial difficulties in utilising and managing information distributed across multiple PIM tools throughout the scholarly workflow. Information fragmentation has long been recognised as a central challenge in PIM research, with prior work repeatedly calling for greater unification strategies [10, 16]. In the scholarly context, our findings highlight the need to bridge PIM tools used across different types of scholarly

activities and to integrate their information management capabilities, thereby enabling more coherent and unified management of scholarly information across the entire workflow.

Despite the insights yielded by this study, the current research design and process exhibit several limitations that point to directions for future work. First, some participants, particularly those who already possessed well-articulated personal scholarly workflows, reported confusion when asked to select terms to build their own scholarly workflow models in the user survey. For such participants, imposing predefined vocabulary may obscure rather than clarify their practices. Future extensions of our study will therefore allow greater flexibility, enabling participants to skip certain questionnaire sections. Second, the participant sample for the guided tour was composed largely of PhD candidates representing early-career researchers. In future work, we seek to recruit a more diverse participant pool for the guided tour, including researchers at later career stages, such as research assistants and professors. Third, although our sample included participants from multiple disciplines, we did not systematically analyse differences in PIM practices across disciplines. Preliminary observations suggest that humanities scholars tend to rely more on general and even paper-based tools, such as file folders, Word and printed articles, whereas science and engineering scholars more readily adopt digital and niche information tools, such as Paperlib and Obsidian. In future research, we also plan to examine how disciplinary background and factors such as technical literacy influence scholarly workflows and scholarly PIM practices.

6 Conclusion

The primary contribution of this paper is to take a first step towards addressing a gap in scholarly PIM literature by introducing the concept of *scholarly workflows* into the domain of scholarly PIM studies. This integration opens new possibilities and offers a fresh and valuable perspective, enriching the scholarly PIM discourse. By framing scholarly PIM within the broader context of scholarly workflows, researchers are empowered to adopt a top-down, holistic perspective that reveals how scholarly PIM activities are embedded across the entire scholarly workflow. This conceptual advancement not only deepens our understanding of scholarly PIM practices from a holistic, workflow-oriented perspective but also opens up new avenues for designing tools and systems that more effectively support scholars' evolving information management and human-information interaction needs.

Acknowledgments

The work of Qi Xu is supported by the China Scholarship Council (CSC) through a doctoral scholarship (Grant No. 202408330076).

References

[1] Mashael Al-Omar and Andrew Martin Cox. 2016. Scholars' Research-related Personal Information Collections: A Study of Education and Health Researchers in a Kuwaiti University. *Aslib Journal of Information Management* 68, 2 (March 2016), 155–173. doi:10.1108/AJIM-04-2015-0069

[2] Smiljana Antonijević. 2015. Workflows of Digital Scholars. In *Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production.* Palgrave Macmillan US, 37–72. doi:10.1057/9781137484185_3

[3] Smiljana Antonijević and Ellysa Stern Cahoy. 2014. Personal Library Curation: An Ethnographic Study of Scholars' Information Practices. *portal: Libraries and the Academy* 14, 2 (April 2014), 287–306. doi:10.1353/pla.2014.0010

[4] Ofer Bergman and Steve Whittaker. 2016. *The Science of Managing our Digital Stuff.* The MIT Press.

[5] Jeroen Bosman and Bianca Kramer. 2015. 101 Innovations in Scholarly Communication: How Researchers Are Getting to Grip with the Myriad of New Tools. *Impact of Social Sciences Blog* (2015). https://blogs.lse.ac.uk/impactofsocialsciences/2015/11/11/101-innovations-in-scholarly-communication/

[6] Kaila Bussert, Kathy Chiang, and Kornelia Tancheva. 2011. Personal Management of Scholarly Information. In *Scholarly Practice, Participatory Design and the eXtensible Catalog*. Association of College & Research Libraries (ACRL), 123–151.

[7] Joel Chan, Xin Qian, Katrina Fenlon, and Wayne Lutters. 2020. Where the Rubber Meets the Road: Identifying Integration Points for Semantic Publishing in Existing Scholarly Practice. In *JCDL Workshop on Conceptual Modeling*. Wuhan, China. https://github.com/sig-cm/JCDL-2020/blob/master/JCDL_Where_the_rubber_meets_the_road_2020-6-28-FINAL.pdf

[8] Abdus Sattar Chaudhry and Bibi M. Alajmi. 2022. Personal Information Management Practices: How Scientists Find and Organize Information. *Global Knowledge, Memory and Communication* 73, 6/7 (2022), 757–774. doi:10.1108/GKMC-04-2022-0082

[9] Danielle Cooper and Oya Rieger. 2018. *Scholars ARE Collectors: A Proposal for Rethinking Research Support.* Technical Report. Ithaka S+R. doi:10.18665/sr.310702

[10] Alan Dix. 2024. The Future of PIM: Pragmatics and Potential. *Human–Computer Interaction* 41, 2 (June 2024), 1–28. doi:10.1080/07370024.2024.2356155

[11] Nancy Falciani-White. 2017. Information Behaviors of Elite Scholars in the Context of Academic Practice. *Journal of Documentation* 73, 5 (October 2017), 953–973. doi:10.1108/JD-02-2017-0028

[12] Han L Han, Junhang Yu, Raphael Bournet, Alexandre Ciorascu, Wendy E Mackay, and Michel Beaudouin-Lafon. 2022. Passages: Interacting with Text Across Documents. In *Proceedings of CHI 2022, Conference on Human Factors in Computing Systems*. New Orleans, USA, 1–17. doi:10.1145/3491102.3502052

[13] Sharon Ince, Christopher Hoadley, and Paul A. Kirschner. 2020. Research Workflow Skills for Education Doctoral Students and Postdocs: A Qualitative Study. *The Journal of Academic Librarianship* 46, 5 (September 2020). doi:10.1016/j.acalib.2020.102172

[14] William Jones. 2014. *The Future of Personal Information Management, Part 2: Transforming Technologies to Manage Our Information.* Springer International Publishing, Cham. doi:10.1007/978-3-031-02329-3

[15] William Jones. 2015. *The Future of Personal Information Management Part 3: Building a Better World with Our Information.* Springer International Publishing, Cham. doi:10.1007/978-3-031-02295-1

[16] David R. Karger and William Jones. 2006. Data Unification in Personal Information Management. *Commun. ACM* 49, 1 (January 2006), 77–82. doi:10.1145/1107458.1107496

[17] Jiali Liu and James Eagan. 2021. ADQDA: A Cross-Device Affinity Diagramming Tool for Fluid and Holistic Qualitative Data Analysis. *Proceedings of the ACM on Human-Computer Interaction* 5, ISS (November 2021), 1–19. doi:10.1145/3488534

[18] John S Morabito and Joel Chan. 2021. Managing Context during Scholarly Knowledge Synthesis: Process Patterns and System Mechanics. In *Proceedings of C&C 2021, 13th Conference on Creativity and Cognition.* ACM, Virtual Event Italy, 1–5. doi:10.1145/3450741.3465244

[19] Williams Ezinwa Nwagwu and Antonia Bernadette Donkor. 2021. Personal Information Management Behaviors of University Faculty - Aspects of Print versus Electronic. *Libri* 71, 2 (June 2021), 183–202. doi:10.1515/libri-2020-0041

[20] Carole L. Palmer, Lauren C. Teffeau, and Carrie M. Pirmann. 2009. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development.* OCLC Research, Dublin, Ohio.

[21] Peter Pirolli and Stuart Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis. In *Proceedings of the International Conference on Intelligence Analysis*, Vol. 5. McLean, USA, 2–4.

[22] Xin Qian, Matt J Erhart, Aniket Kittur, Wayne G Lutters, and Joel Chan. 2019. Beyond iTunes for Papers: Redefining the Unit of Interaction in Literature Review Tools. In *Companion Publication of CSCW 2019, International Conference on Computer Supported Cooperative Work and Social Computing*. Austin, USA, 341–346. doi:10.1145/3311957.3359455

[23] Daniel M. Russell. 2003. Learning to See, Seeing to Learn: Visual Aspects of Sensemaking. In *Proceedings of Electronic Imaging 2003*. Santa Clara, USA. doi:10.1117/12.501132

[24] Ciaran B. Trace and Unmil P. Karadkar. 2017. Information Management in the Humanities: Scholarly Processes, Tools, and the Construction of Personal Collections. *Journal of the Association for Information Science and Technology* 68, 2 (February 2017), 491–507. doi:10.1002/asi.23678

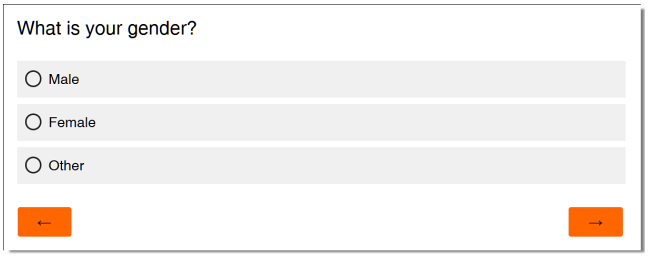
[25] Steven Weiland. 2018. The Scholarly Workflow in the Digital Age: What Do We Know? What Should We Do?. In *Proceedings of the Charleston Library Conference*. doi:10.5703/1288284316687

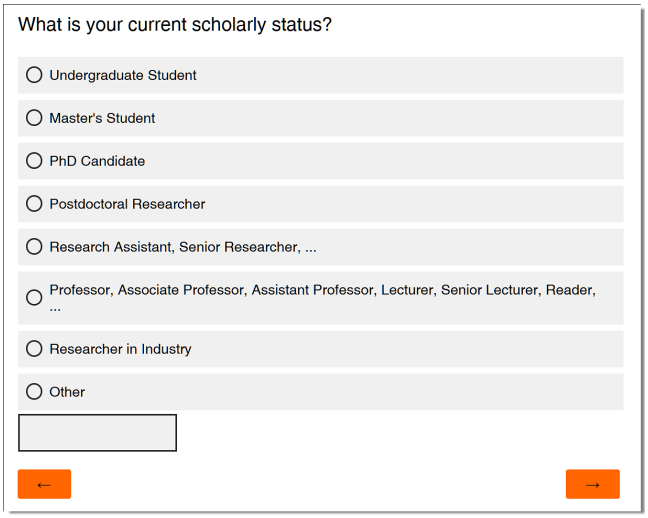
[26] Shazia Yasmin, Nosheen Fatima Warraich, and Irfan Ali. 2019. Personal Digital Information Management Practices of Engineering Faculty: Finding, Organizing, and Re-finding Information. *PJIM&L* 21 (December 2019), 88–103. https://pjiml.pu.edu.pk/jo/index.php/pjiml/article/view/55

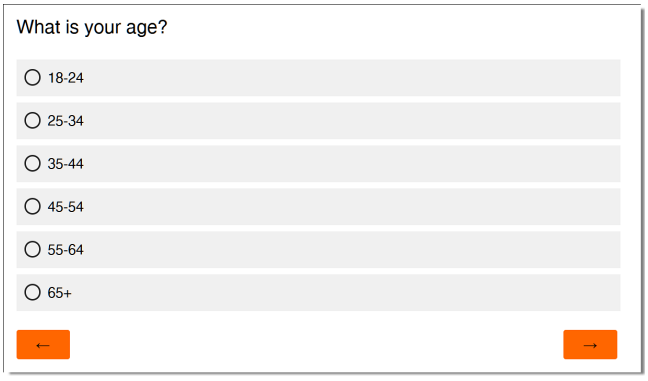
A User Survey Questionnaire

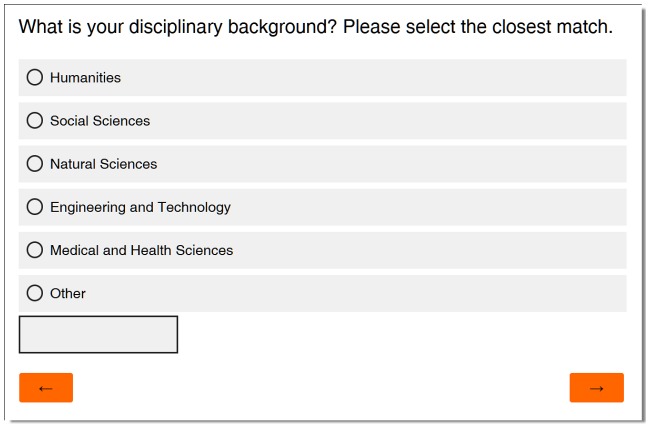
Note that the full, anonymised dataset containing the user study and guided tour interview data, along with all the questions, is available at https://doi.org/10.5281/zenodo.18742313.

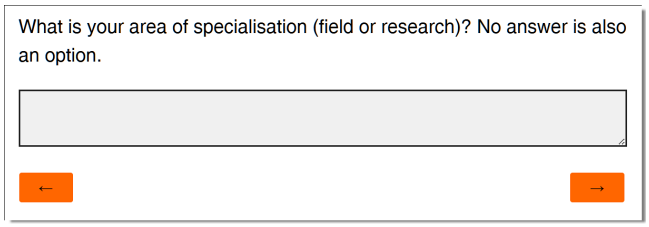
A.1 Example of Questions in First Section

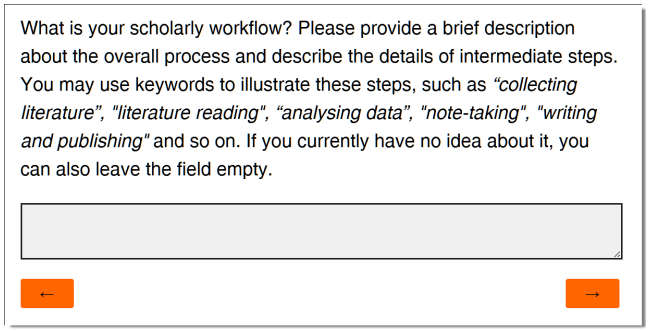








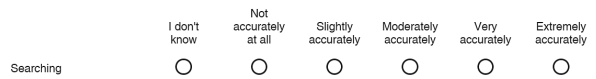




A.2 Example of Questions in Second Section

Below are the scholarly activities defined by Palmer's (2009). Please rate each of Palmer's scholarly activities based on how accurately they describe the different parts of your personal scholarly workflow.

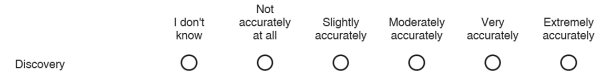
- **Searching**
  - *direct searching:* use familiar keywords and terms in various databases and online platforms to locate specific information.
  - *chaining*: trace bibliographic references in scholarly works to get relevant materials.
  - *browsing*: look through a body of assembled or accessible information.
  - *probing*: seek relevant information outside the area of expertise.
  - *accessing*: directly and efficiently access to identified materials.



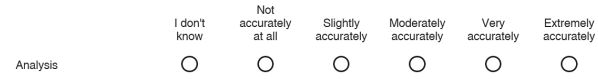
- **Collecting**
  - *gathering*: build collections to maintain long-term accessibility and convenience.

Below are the scholarly activities defined by Bosman and Kramer's (2015). Please rate each of Bosman and Kramer's scholarly activities based on how accurately they describe the different parts of your personal scholarly workflow.

- **Discovery**
  - search literature/data etc.
  - get access to literature etc.
  - get alerts/recommendations
  - read/view/annotate

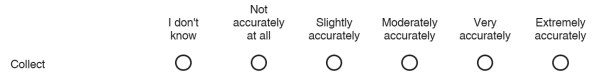


- **Analysis**
  - analyse data/text
  - share notebooks/protocols/workflows

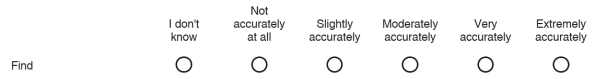


Below are the scholarly activities defined by Antonijević's (2015). Please rate each of Antonijević's scholarly activities based on how accurately they describe the different parts of your personal scholarly workflow.

- **Collect**
  - gathering research data and materials

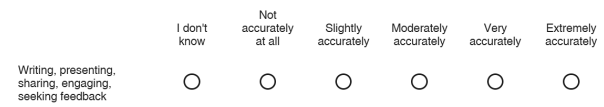


- **Find**
  - finding research data and materials

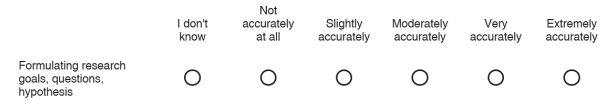


Below are the scholarly activities defined by Cooper and Rieger's (2018). Please rate Couper and Rieger's scholarly activities based on how accurately they describe the different parts of your personal scholarly workflow.

- **Writing, presenting, sharing, engaging, seeking feedback**

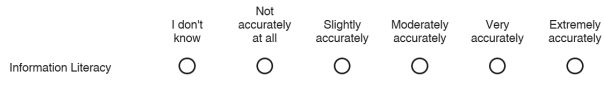


- **Formulating research goals, questions, hypothesis**

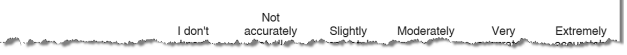


Below are the scholarly activities defined by Ince's (2020). Please rate Ince's scholarly activities based on how accurately they describe the different parts of your personal scholarly workflow.

- **Information Literacy**
  - finding and evaluating the right information
  - library search, alerting, discovery, relevance judgments



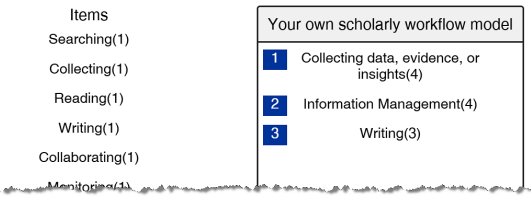
- **Information Management**
  - organising and storing information
  - bibliographic management, personal repositories



A.3 Example of Questions in Third Section

Here are your ratings of scholarly activities from the five existing scholarly workflow models, using a scale from 1 to 5, where 5 indicates that the vocabulary extremely accurately describes an intermediate step of your scholarly workflow. If you selected 'I don't know' in the previous options, it will be displayed as 0.

Now, using the vocabularies from these models that you believe best describe the intermediate steps of your scholarly workflow, please build up your own scholarly workflow model by dragging and dropping the terms from the left into the box on the right. Please choose no more than 6 activities that, taken together, can broadly cover your scholarly workflow and represent what you consider most essential. When you place the terms from the left into the box on the right, they will be automatically arranged in order. From top to bottom, this represents the temporal direction of your scholarly workflow. Since scholarly workflows are not always perfectly linear, the order does not need to be exact—just an approximate sequence.



A.4 Example of Questions in Fourth Section

What types of scholarly information do you encounter throughout scholarly activity "Writing(3)"? Common types of scholarly information include related work (PDFs), experimental data, emails, notes, drafts, annotations and so on.

What information management tools do you use in scholarly activity "Writing(3)"? Common types of scholarly information management tools include file manager, email manager, citation management tool(e.g. Zotero, Mendeley, EndNote), qualitative data analysis tools(e.g. NVivo) and so on.

In scholarly activity "Writing(3)", how do you manage related scholarly information? What do you think are the shortcomings or limitations of your current approach? You may try to answer these two questions from the perspectives of information finding, refinding, keeping, organising

What types of scholarly information do you encounter throughout scholarly activity "Collecting data, evidence, or insights(4)"? Common types of scholarly information include related work (PDFs), experimental data, emails, notes, drafts, annotations and so on.

What information management tools do you use in scholarly activity "Collecting data, evidence, or insights(4)"? Common types of scholarly information management tools include file manager, email manager, citation management tool(e.g. Zotero, Mendeley, EndNote), qualitative data analysis tools(e.g. NVivo) and so on.

In scholarly activity "Collecting data, evidence, or insights(4)", how do you manage related scholarly information? What do you think are the shortcomings or limitations of your current approach? You may try to answer these two questions from the perspectives of information finding

B Example Screenshots from the Guided Tour Interviews

B.1 PIM Tools for Collecting Scholarly Activities

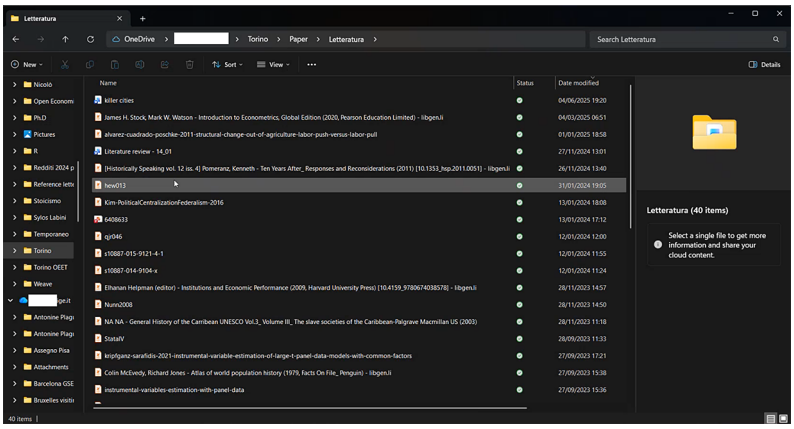


Figure 1: File Explorer (Windows file manager)

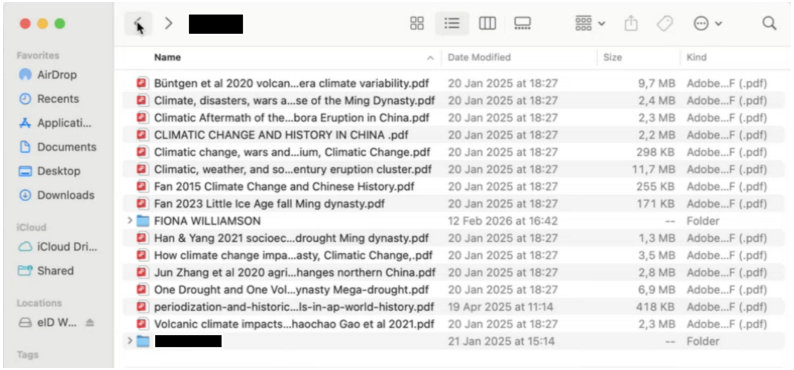


Figure 2: Finder (MacOS file manager)

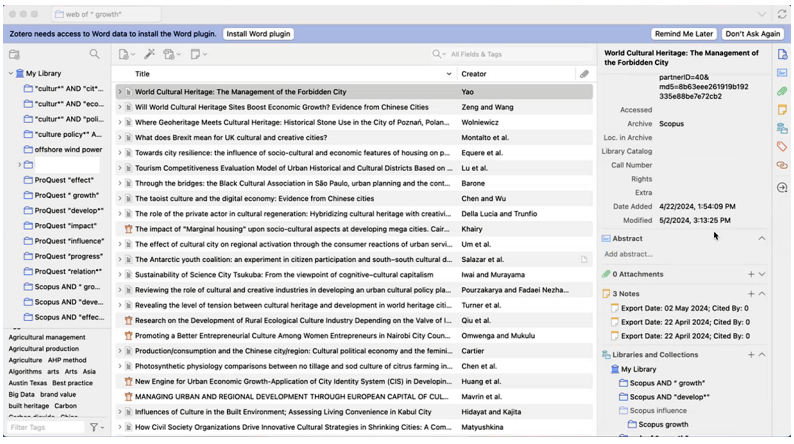


Figure 3: Zotero citation management tool

B.2 PIM Tools for Curating Scholarly Activities - Annotations

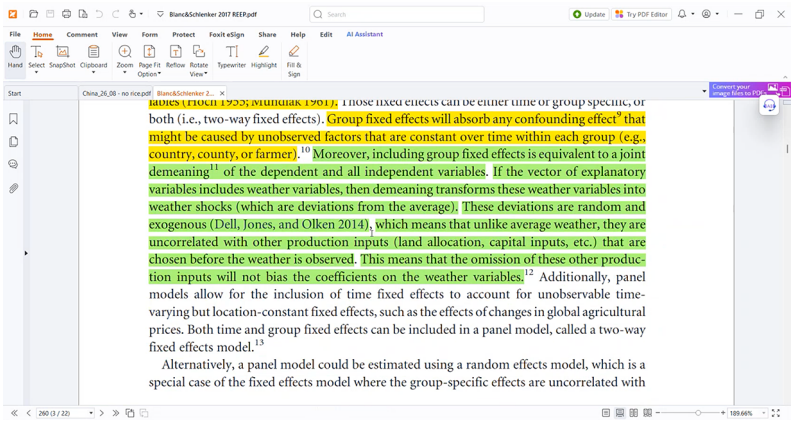


Figure 4: Foxit PDF Viewer standalone desktop application

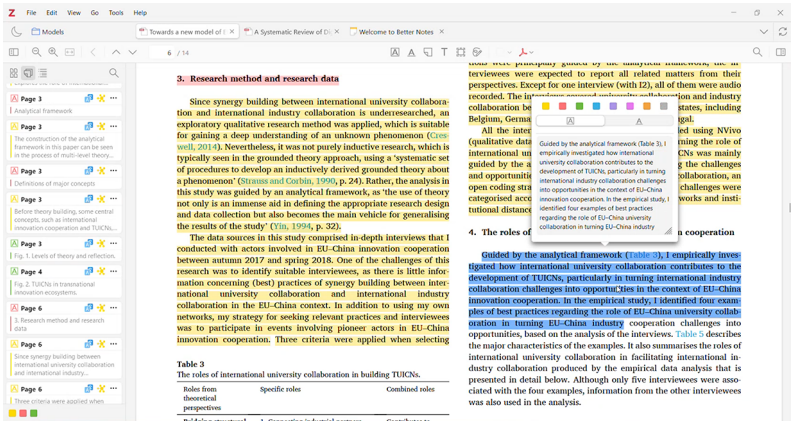


Figure 5: PDF Viewer embedded in Zotero CMT

B.3 PIM Tools for Curating Scholarly Activities - Notes

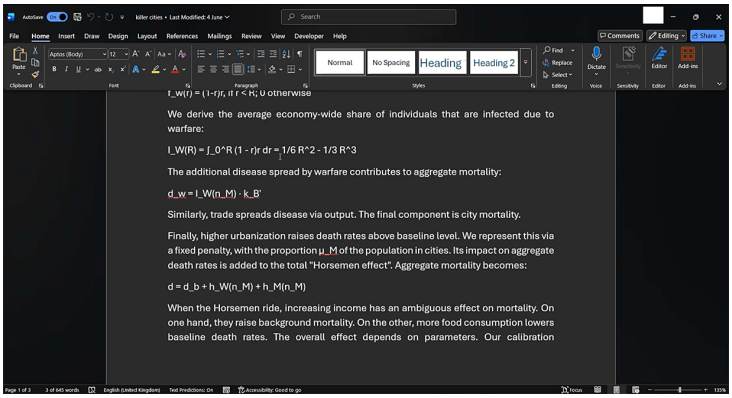


Figure 6: Word - Note-taking through paraphrasing

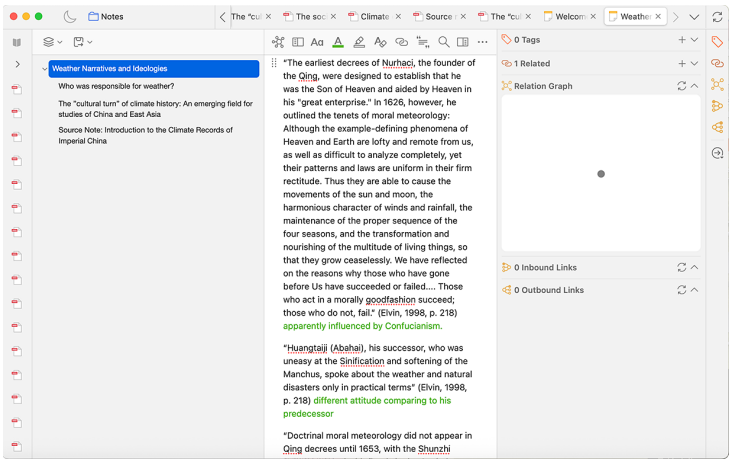


Figure 7: Zotero's note module - Note-taking through copy/paste and links

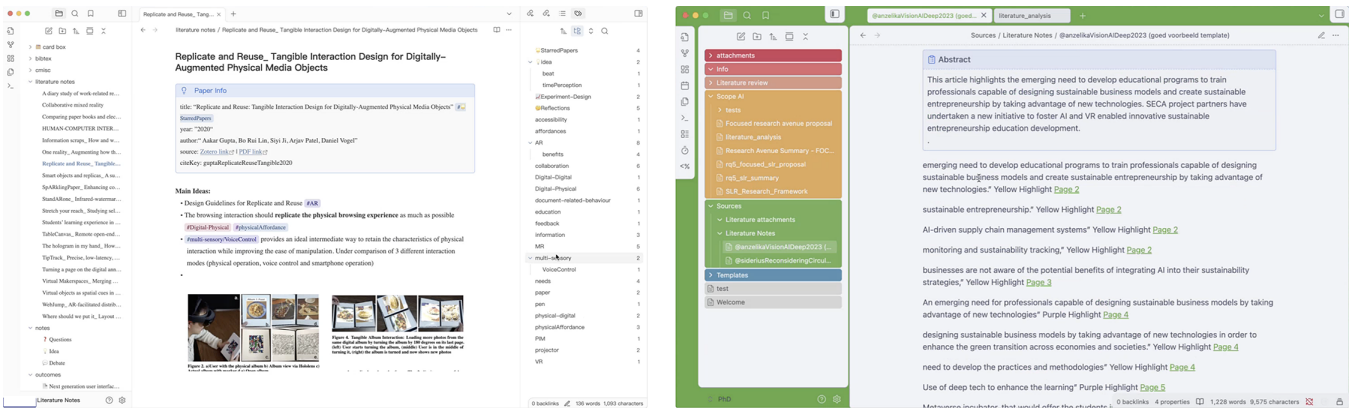


Figure 8: Obsidian - Note-taking through paraphrasing, copy/paste, screenshots and links

B.4 Sensemaking PIM Tool Features for Curating Scholarly Activities

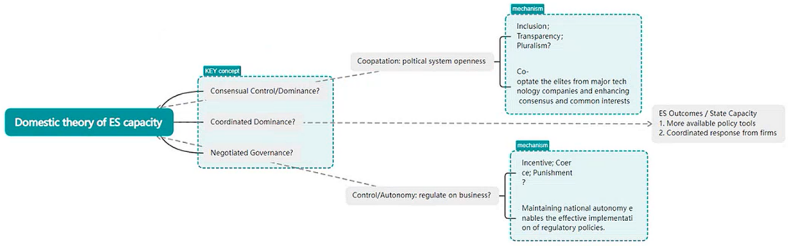


Figure 9: Mind Map

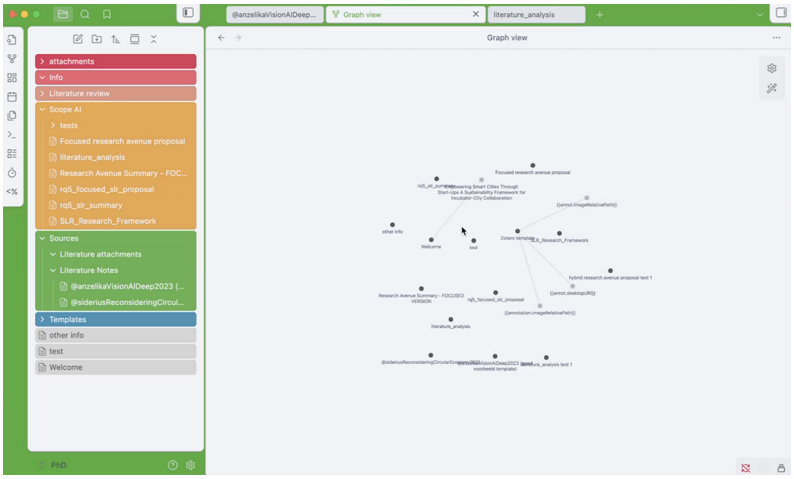


Figure 10: Node-Link Diagram