

An Empirical Study on Personal Information Management Practices Across Scholarly Workflows

Qi Xu and Beat Signer

Web & Information Systems Engineering Lab, Vrije Universiteit Brussel
qi.xu@vub.be, bsigner@vub.be

Background

Scholarly Personal Information Management (Scholarly PIM): Scholarly PIM refers to how scholars collect, organise, maintain and make sense of research-related information throughout their work.

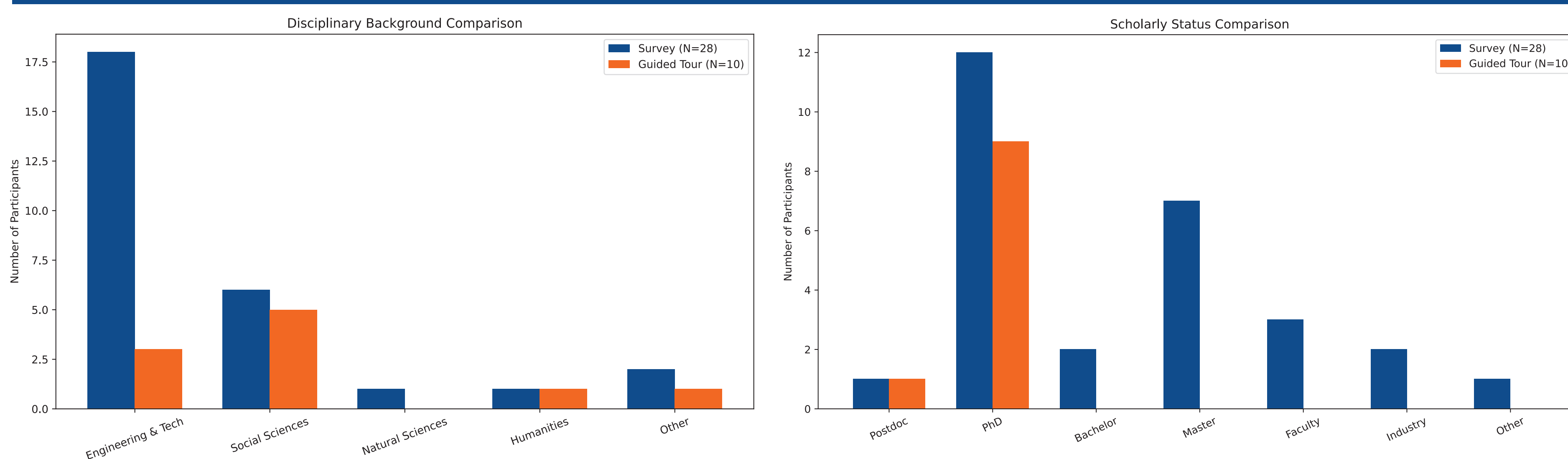
Scholarly Workflow: Scholarly workflows refer to the series of interconnected scholarly activities that span the research lifecycle, from finding, storing and analysing to writing, sharing and publishing.

Research Gap: Current research on scholarly PIM has yet to fully incorporate this concept into its analytical framework, offering limited understanding of how PIM practices operate within and across activities, as well as how these practices vary or evolve across stages.

Research Objective: Our work aims to investigate scholarly PIM practices across workflows, identify their limitations, and inform the design of better support tools.

Time	Discipline	Author(s)	Scholarly Workflow (Scholarly Activities)
2009	Scholars across fields	Palmer et al.	Searching (direct searching, chaining, browsing, probing, accessing), collecting (gathering, organising), reading (scanning, assessing, rereading), writing (assembling, co-authoring, disseminating), collaborating (coordinating, networking, consulting), monitoring, note-taking, translating, data practices
2015	Humanities	Antonijević	Collect, find, analyse, write, communicate, organise, annotate, cite, reflect, archive and share
2015	Scholars across fields	Bosman and Kramer	Discovery, analysis, writing, publication, outreach and assessment
2018	Scholars across fields	Cooper and Reiger	Writing, presenting, sharing, engaging, seeking feedback; formulating research goals, questions, hypothesis; collaborating or networking; conducting a literature review; grant writing and fund raising; selecting data gathering and experimental methods; collecting data, evidence, or insights; interpreting findings and testing hypotheses; analysing data, evidence, or insights; and assessing findings and conducting quality control
2020	Social Science	Ince et al.	Information literacy, information management, knowledge management and scholarly communication

Methods



User survey: We conducted a user survey enabling scholars to construct their own workflow models and collected self-reported data on their personal information management practices across different stages of their workflows.

Guided tour study: We carried out a guided tour study to obtain in-depth insights into scholars' real-world PIM practices by observing and discussing how they organise and use information in their actual working environments.

Findings

Collecting Scholarly Activities: Collecting scholarly activities focus on gathering raw primary materials (e.g. experimental or survey data) and secondary materials (e.g. papers, books or web content) by searching external sources, extracting relevant information and storing it locally or remotely for future retrieval.

Curating Scholarly Activities: Curating scholarly activities centre on interpreting, analysing and processing raw material obtained during collecting activities to derive meaningful insights.

Creating Scholarly Activities: In creating scholarly activities, scholars transform material gathered during collecting and processed during curating into mature research output.

Cross-cutting Scholarly Activities: Some scholarly activities, such as 'Translating' span multiple types and correspond to "cross-cutting" scholarly activities, which are commonly co-present alongside many other scholarly activities.

Type	Collecting	Curating	Creating
Scholarly Activities	Searching [5] Collecting [5] Monitoring [5] Collect [1] Find [1] Discovery [2] Collecting data, evidence or insights [3] Information Literacy [4]	Reading [5] Analyse [1] Annotate [1] Analysis [2] Analysing data, evidence or insights [3] Knowledge Management [4]	Writing [5] Write [1] Cite [1] Writing [2] Publication [2] Outreach [2] Writing, presenting, sharing, engaging, seeking feedback [3] Grant writing and fund raising [3] Interpreting findings and testing hypotheses [3]
	Formulating research goals [3] Selecting data gathering and experimental methods [3]		
	Notetaking [5], Translating [5], Data Practices [5], Collaborating [5], Organise [1], Reflect [1], Archive [1], Communicate [1], Share [1], Assessment [2], Conducting a literature review [3], Assessing findings and conducting quality control [3], Collaborating or networking [3], Information Management [4], Scholarly Communication [4]		

- [1] Smiljana Antonijević. Workflows of Digital Scholars. In *Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production*. Palgrave Macmillan US, 2015.
- [2] Jeroen Bosman and Bianca Kramer. 101 Innovations in Scholarly Communication: How Researchers Are Getting to Grip with the Myriad of New Tools. *Impact of Social Sciences Blog*, 2015.
- [3] Danielle Cooper and Oya Rieger. *Scholars ARE Collectors: A Proposal for Re-thinking Research Support*. Technical report, Ithaka S+R, 2018.
- [4] Sharon Ince et al. *Research Workflow Skills for Education Doctoral Students and Postdocs: A Qualitative Study*. *The Journal of Academic Librarianship*, 46(5), 2020.
- [5] Carole L. Palmer et al. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. OCLC Research, Dublin, Ohio, 2009.

• In the **collecting stage**, scholarly information is managed and interacted with at the **file level**.

• In the **curating stage**, scholarly information is managed and interacted with at the **excerpt level**.

• Current excerpt-level PIM strategies provide **limited support** for efficiently **preserving contextual information** associated with excerpts.

PIM Practices Related to Collecting Scholarly Activities

- Keeping:**
- Objects:
 - Primary materials, like research data (e.g. CSV, MP4)
 - Secondary materials, like academic papers (PDFs)
 - Methods:
 - Put into File Managers (e.g. Finder or File Explorer)
 - Put into Citation Management Tools (e.g. Zotero)
- Organising:**
- Using "folder" or "tag" structures
 - Unsynchronised structures
- Maintaining:**
- Current tools provide basic maintenance functions
 - CMTs additionally offer basic duplication detection

PIM Practices Related to Curating Scholarly Activities

- Annotations**
- Keeping:** excerpts are retained as highlights within original files (e.g. PDFs, plain text or images)
- Organising:** little to no support, embedded in source documents
- Maintaining:**
- basic maintenance supported (update and delete)
 - local context preserved (in-document context)
 - limited context scope (external context)

- Notes**
- Keeping:** excerpts are kept in stand-alone notes via copy/paste, screenshots, links or paraphrasing
- Organising:** notes serve as grouping units for organising excerpts
- Maintaining:** context loss (original content, background & provenance often missing)
- copy/paste & screenshots: lose background & provenance
 - paraphrasing: lose almost all context
 - links: preserve context, but hard to use and limited (mostly PDF)

- Codes**
- Keeping:** excerpts are kept in specific qualitative analysis tools, like NVivo
- Organising:** codes serve as grouping units for organising excerpts
- Maintaining:** strong context preservation, but heavyweight and cumbersome

