I (Don't) Know What You Did Last Summer: A Framework for Ubiquitous Research Preservation

Passant Elagroudy
German Research Centre for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
LMU Munich
Munich, Germany
passant.elagroudy@gmail.com

Pascal Knierim
University of the Bundeswehr Munich
Munich, Germany
pascal.knierim@unibw.de

Paweł W. Woźniak
Chalmers University of Technology
Gothenburg, Sweden
pawelw@chalmers.se

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Sebastian S. Feger
LMU Munich
Munich, Germany
sebastian.feger@ifi.lmu.de

		Initiative	
		User-initiated	**Machine-initiated**
Resource Awareness	**Conscious**	User decides to explicitly document pieces of the information. **E.g.:** *uploading to repositories/ writing meeting minutes in a notebook*	System proposes starting the capture highlighting impact + provides user with control over duration/ captured content **E.g.:** *prompts to save information – prompts to record online meetings*
	Unaware	The user initiates the capture + controls duration but it happens in the background. **E.g.:** *time sheet automated logging over project tasks.*	Continuous automated capturing over undefined periods **E.g.:** *automated recording from the institute of screen activities*.

Figure 1: A two-dimensional model explaining user interactions in Ubiquitous Research Preservation (URP) systems (Model 1).

ABSTRACT

Research preservation is a pillar for knowledge transfer, science reproducibility and saving time by reusing existing resources. However, human compliance with efficient capturing strategies is a key barrier to creating complete scientific repositories. To circumvent this issue, we introduce the term: *Ubiquitous Research Preservation* (URP), describing automated knowledge capturing and retrieval in computational science. We also propose a framework composed of three models for designing URP systems (URPS) to 1) understand users' interaction and data governance, 2) propose technical pipelines for data management, and 3) understand users' sharing practices. Our work is a theoretical reflection on our past experiences in designing URPS. We plan future evaluation by using the framework to analyze existing URPS. We expect a positive impact from using URPS on researchers' sense-making and ability to share findings and resources. Our framework is a checklist for design decisions needed to build successful URPS.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; **Ambient intelligence**.

KEYWORDS

Ubiquitous Research Preservation, Reproducibility, Sensemaking, Training, Computational Science

ACM Reference Format:
Passant Elagroudy, Pascal Knierim, Paweł W. Woźniak, Albrecht Schmidt, and Sebastian S. Feger. 2023. I (Don't) Know What You Did Last Summer: A Framework for Ubiquitous Research Preservation. In *Extended Abstracts*

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585754

of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3544549.3585754

1 INTRODUCTION

Preserving and sharing knowledge is an integral part of scientific research. Prior work has shown that ubiquitous logging of knowledge workers' routines can aid in performance assessments, spurring a sense of productivity, and providing an opportunity for reflection and process optimization [6]. Additionally, such logs act as memory prosthesis reminding the worker of their workflow, saving time and energy spent in understanding and/or rebuilding existing systems (e.g. [15, 19]). Moreover, they can be used to transfer sparse knowledge to others forming an augmented educational system (e.g. [5, 20, 21]).

In this paper, we focus on the context of *knowledge preservation in computational and data-driven sciences* (KPS), known as the third and fourth paradigm of science [18]. The aforementioned benefits apply to researchers as they are considered knowledge workers. The special nature of research as a context for knowledge preservation is predominant in four scenarios: 1) longitudinal work, 2) documenting unyielding approaches, 3) data protection laws, and 4) research reproducibility. Firstly, research projects are often revisited in a longitudinal manner due to the longer publishing cycles and/or passed to students to extend them. However, interpreting larger code bases over longer periods of time is a major problem [27]. Secondly, positive results are usually favoured in scientific publications over unyielding attempts. Thus, logs can offer alternative archives to save future resources from repeating such attempts and/or simplify building upon them. Thirdly, research data is often strictly governed by data protection laws from several albeit competing stakeholders such as research institutes, funding agencies, local laws in hosting countries. Fourthly, there is a trend calling for providing supplementary material to support research reproducibility (e.g. reproducible research data movement to provide data and code of experiments [8], and [9, 25]).

Automated preservation is handy in the context of computational and data-driven sciences as prior work recommended documenting the workflow of scientific processes rather than only supplementary data [1, 27]. Additionally, Tanaka et al.[29] also identified problems in documentation-centric tasks such as: organizing growing number of resources, and recalling the relevant documents to particular work situations. This can be partially circumvented by automated data captures. Feger et al. [11] showed that values like extensive communication and uncertainty about the ownership of resources are essential to consider while designing automated systems preserving research. They also proved a user need for automation of preservation.

While there is rich literature about design models for personal informatics systems (e.g. [4, 17]), there is sparsity in ones specifically targeting the unique aforementioned challenges of KPS. There are also several on-going efforts to document best practices for archiving experiments and scientific work (e.g. [27] and digital collections guidelines by University of Illianos[1]). However, they aim to regulate the researcher's practices rather than offer guidelines to design automated systems. To this end, we complement prior work and present two contributions in this position paper:

(1) We coin and define the term *"Ubiquitous Research Preservation"* (URP).
(2) We propose three models for designing URP systems covering the design cycle. The first models the user governance and interaction with the digital artefacts. The second covers a technical pipelines for building URP systems, i.e. the data management. The third and last model is about data sharing. The three models together form our proposed framework for designing systems for research preservation.

Our work summarizes our experience in designing URP systems (such as [6, 11–14, 16]). The proposed framework is not meant as a comprehensive tool fitting any design problem. On the contrary, this work is a starting point to ignite discussions about creating frameworks for the special area of research preservation. It aids designers in building tools to support on going endeavours for research preservation and replication.

2 BACKGROUND AND RELATED WORK

In this section, we first provide an overview from prior work about features identified as essential for successful automation of research preservation. Current literature hints towards the value of research preservation for: scientist's own reflection; supporting training; and the reproducibility of research. Thus, we shortly provide examples for those use cases under the lens of preservation in computation and data-based sciences.

2.1 Features of Automated Systems for Research Preservation

The design and study of tools that support researchers in documenting knowledge has a long tradition in Human-Computer Interaction (HCI) and Computer-Supported Cooperative Work (CSCW). As digital technologies became prevalent, research focused on enhancing or replacing traditional, physical lab notebooks with electronic ones. Studying the use of a hybrid laboratory notebook, Tabard et al. [28] found that *"when trying to complete a task, users clearly do not want to focus on the process of capturing information"*. Yet, they also noted that automated mechanisms can be intrusive and that users need to be in control of the recording and sharing. Oleksik et al. [26] reported on their observational study on electronic lab notebooks (ELN). They found that the flexibility of digital media can lead to much less precision during experiment recording and that "freezing" parts of the record might be necessary. The authors also stressed that *"ELN environments need to incorporate automatic or semi- automatic features that are supported by sophisticated technologies [...]."* However, it is generally not very clear how automated recording and need for control can play together.

Kery et al. [22] reported on a study of literate programming tools conducted with data scientists. They asked scientists to think about *"a magical perfect record of every analysis run [...]"*. Participants created queries where they *"referred to many kinds of contextual details, including libraries used, output, plots, data sources, [...]."* Participants described their inability to find prior analyses and illustrated consequences. The authors found that in literate programming tools,

[1]Last accessed from: https://guides.library.illinois.edu/c.php?g=465430 on 2023-01-18

"version control is currently poor enough that records of prior iterations often do not exist". These results also conform with the findings from another relevant knowledge domain, that is documentation work. In their experiments, Tanaka et al.[29] found a need for intuitive ways to efficiently recall the location of deliverable files based on the work situation.

Feger et al. [12] found that that motivation is a key element in technology design and that even gameful interactions can create motivation in highly skilled environments. Additionally, Feger et al. [11] recommended the usage of domain-specific templates to support analysts in automating preservation and verification of their research. Prior work also highlights conflicting factors like privacy and control.

2.2 Benefits of Research Preservation

2.2.1 Aiding Recall, Sensemaking and Reflection. Tabard et al. [28] report on their studies of a hybrid laboratory notebook that they used as a technology probe with biologists. They illustrate the importance of reflection in the scientific process and highlight how access to preserved information can support researchers in their reflection. They further stress that the presentation format greatly impacts this process. Redundant information appearing again and again help *"scientists understand how their thoughts have evolved over time."* Thus, the availability and information review can help to gain new insights and support creativity.

2.2.2 Impact on Research Reproducibility. The effort needed to preserve and share experimental data has been identified as one of the core contributors to the observed reproducibility challenge [2]. And even though preservation and sharing of digital artifacts is rather easy, reproducibility of computational studies remain a concern [9, 27]. For example, Echtler et al. [9] show that only 21 out of 1027 papers in "CHI 2016"[2] released their source code although the papers that used open source code is five-fold that number. While current data and code repositories provide a powerful means to preserve and share work, they capture only snapshots and do not record entire processes. Recording and potentially sharing work as a sum of changes can become feasible, if we implement automated recording strategies. In addition, this will allow distribution of further context information. Preserving for example meta-data about the computational setup like runtime configurations and architecture, we can provide information relevant for full reproducibility that are not commonly available.

2.2.3 Supporting Education and Training. Oleksik et al. [26] studied use of an ELN in a teaching research organization. They illustrate researchers' desire to complement guidance by their peers and supervisors with a system that recommends relevant resources during experiment design and execution. Basis for such recommendations could be the organization's documented and preserved experiments. Further opening preserved resources can profit education and enable replication studies from which researchers can learn [30].

[2]The ACM CHI Conference on Human Factors in Computing Systems is the premier international conference of Human-Computer Interaction (HCI)

3 DEFINITION OF "UBIQUITOUS RESEARCH PRESERVATION"

Preservation in computation science refers to the documentation of computational resources essential in demonstrating evidence; either for internal or external purposes. Those resources include data, data-processing software, results, meta-data that describe the computational setup, and the scientific workflow. Resources' storage must not be predeterminedly limited at the time of recording. **Ubiquitous Research Preservation (URP)** refers to the machine-supported scientific knowledge recording and preservation process of computational workflows. URP technology *initiates* and/or *controls* partial or complete process capturing. URP systems are a unique use case crossing the borders between personal informatics systems (PIS) and memory prostheses (MP). PIS primarily focus on reflection and behavioural change, while MP generally targets memory alterations be it augmentation (like URP systems) or reformations [10].

We contrast the scope of the URP term with the relevant but different *pervasive digitization* [27] term. Pervasive digitization is a bigger umbrella focusing on all processes related to conducting research involving *participants and researchers* alike. This includes creating architectures and systems for continuous data collections from participants for example. On the other hand, URP primarily focuses only on the researcher's narrative and on documenting their workflows to assist them and their future successors.

4 FRAMEWORK DESIGN

We borrow from the rich literature about designing personal informatics systems and memory prostheses to propose a technical pipeline for implementing such systems. Models in prior work typically fall under *two* categories: (1) describing user processes and interactions, and (2) describing steps for technical implementation. Similarly, we also propose below *two models* representing each category in addition to a model for understanding sharing practices of the data. The three models together form our framework.

4.1 Model 1: User-Centric Interaction Model

We present a two-dimensional model for explaining the user's interaction with the system to maintain data governance and control. Figure 1 provides an overview.

(1) Initiative: Nominates the entity responsible for initiating a preservation process.
  - **User-Initiated:** The researcher initiates and controls the preservation process on-demand.
  - **Machine-Initiated:** The machine initiates and controls processes. Decisions might be based on: workflow knowledge; pre-configured domain rules; and / or user-configured rules.

(2) Resource Awareness: Tackles the researchers' awareness about the selected resources for preservation.
  - **Conscious:** Only those resources are preserved and shared which are actively selected by the user.
  - **Unaware:** The user has no direct control over the resources that are preserved and shared. However, they might have previously set rules for this process.

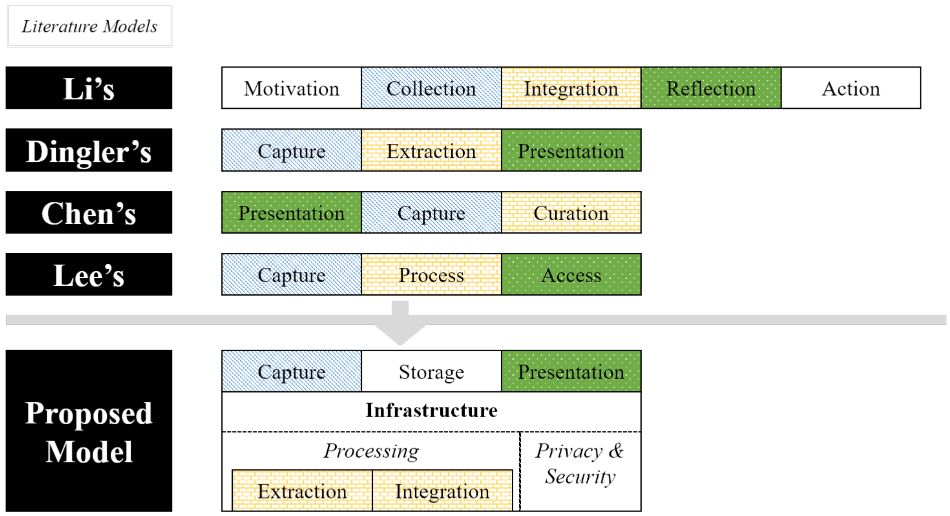

Figure 2: Four-stage pipeline for developing Ubiquitous Research Preservation (URP) Systems (Model 2).

4.2 Model 2: Technical Pipeline for Building URP Systems

We selected four models from the literature commonly used to develop personal informatics and lifelogging systems. Figure 2 summarizes the models and the deduction process leading to the proposed model.

1. (1) **Li's model [24]** The model is for designing personal informatics systems. It has five stages: 1) motivation, where the user decides the reason for lifelogging and what to collect, 2) collection, where the actual data is gathered, 3) integration, where the raw data is transformed into information, 4) reflection, where the user relates the presented information to their behavior, and finally 5) action, where the behavioral change occurs.
2. (2) **Dingler's model [7]** The model is for designing memory augmentation systems. It has three stages: 1) capturing of memory cues, 2) extraction of memory cues, where the selection of a limited amount of cues happens to maximize the quality of memory enhancement, and 3) presentation of memory cues, where the user interacts with the selected cues.
3. (3) **Chen's model [3]** A similar model to Dingler's model [7]. However, the order of the stages is different. It also has three stages: (1) presentation of information similar to Dingler's third stage, (2) capturing of data similar to Dingler's first stage, and (3) retrieval of information referring to the curation and selection process of the cues similar to Dingler's second stage.
4. (4) **Lee's model [23]** The model has three stages as well: (1) capturing the data, (2) processing the data, and (3) accessing the data. The stages also resemble Chen's model but the order is reversed.

The proposed model (see Figure 2) is composed of four layers inspired by the aforementioned literature. The memory cues are collected in the *capture* layer, banked in the *storage* layer, then shown to the user in the *presentation* layer. At each layer, the data is processed, important cues are selected, and all data is protected, forming the *infrastructure* layer.

We extend the existing literature by: 1) consolidating different models into one, 2) separating the "infrastructure" layer so that it is operational in all other layers, and 3) explicitly introducing the "storage" layer. This is particularly relevant to URP systems due to the massive amounts of generated data that pose significant storage and governance challenges. We chose to remove the layers of "motivation" and "action" despite their relevance to the context because the model's focus is the technical aspects rather than the full design cycle.

4.3 Model 3: Understanding Sharing Practices and Stakeholders

Work information captured within URP is a form of memories. We propose a mental model to understand sharing practices inspired by memory prostheses. Within the context of this manuscript, it could aid designers in: (1) negotiating data governance models, (2) identifying privacy and security threats from sharing specific content, and (3) evaluate the truthfulness / completeness of captured information. We discuss *three* dimensions (see Figure 3):

1. (1) **Discoverability** Memories are human captures of life events. *Individual memories* refer to events happening to the person alone without witnesses. *Witnessed memories* refer to events that have bystanders or contributors. Bystanders are passive witnesses like colleagues who accidentally hear work conversations in an office. Contributors are others who are part of the event, like co-authors in a scientific experiment. Individual memories are harder to objectively verify without ubiquitous interventions.
2. (2) **Perspective** *Personal memories* are memories as recalled by an individual even if they happened in a group. *Group memories* are the collective details remembered about an incident by a group of people. Personal memories and group

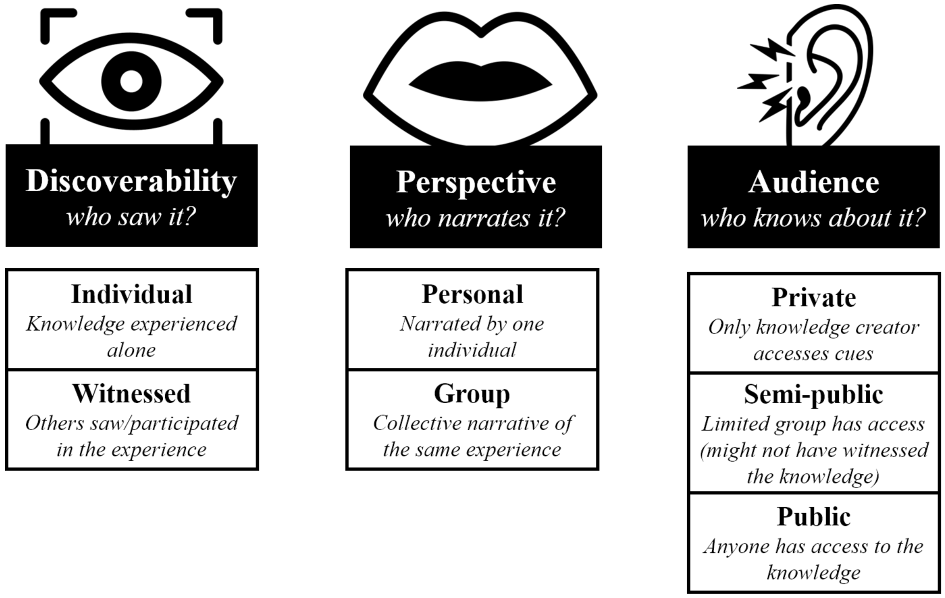

Figure 3: Mental model for sharing practices of knowledge in URP systems (Model 3).

memories might not align about the same incident. For example, narrating the meeting minutes and flow of a meeting might differ between different attendees.

(3) **Audience** Memories and their corresponding documentation could be private (i.e. accessible only to people who experienced it), shared semi-publicly within a limited group of people that did not necessarily experience the incident, or shared publicly within wider audience. In the sharing scenarios, the audience might have access only to the narrative of the sharer. This poses a challenge in verifying the truthfulness and completeness of the shared knowledge. Additionally, there are concerns about giving up leverage of having specialized domain knowledge. An example is a researcher working individually on a project and handing it over to new researchers. The quality of the hand over is limited by the authenticity of the selectively captured then shared information.

5 RESEARCH CHALLENGES, LIMITATIONS AND OPPORTUNITIES

Besides the common challenges of ubiquitous logging systems such as usefulness of collected data and control over captured content, we expand here on some challenges and opportunities special to the context of URP technologies:

Generalizability of templates URP technology will profit from in-depth knowledge of research work in order to efficiently record, discover and present information. The same applies rule-based research preservation. Yet, researchers and especially research domains differ from one another, making the development of general tools difficult. Thus, development and deployment of assistive technology across heterogeneous environments needs to be further researched. Research questions include: How can technology assess researchers' practices, needs and integrate into their workflows? Can we create accessible templates based on learned and confirmed structures? How does technology adapt to scientific novelty and creativity?

Designing storage banks Users are still apprehensive about investing in-storage solutions from near-continuous capturing devices, particularly in unaware machine-initiated scenarios. The captured data has a massive digital footprint. Current solutions do not offer immediate value to users to justify the storage costs. An example is that one-month data of pictorial lifelogs captured the computer screen every 30 seconds consumes approximately 30 Gigabytes[3]. This can easily translate to massive monetary losses to create data banks. Thus, designers are encouraged to invest in compressing the data, particularly in memory augmentation systems.

Reducing reviewing overload The reviewing process of larger datasets remains overwhelming. One technique to review such data sets is parallel reviewing where abstract snapshots of several information cues are presented together (e.g. seeing several scientific papers on a larger screen). Reducing the presentation size of a single cue is crucial in parallel reviewing to avoid overloading the user. We recommend designers identify early on in their systems the sweet spot that masks most of the details while highlighting the salient features. Additionally, smaller grouping units like pages help users in building a mental model of the data to reach meta conclusions such as time spent on a task.

We acknowledge a limitation of our work as the proposed framework is not evaluated yet. However, it is a rather reflective synthesis of the wide expertise of the authors designing relevant systems. We

[3] Average calculation from a one month longitudinal study

plan however to evaluate it in the future qualitatively through a literature review, where we use the proposed framework to categorize the existing systems and using inductive coding to add missing parameters.

6 CONCLUSION

In this paper, we spark discussions on machine-automated preservation in computation-based science. We introduce the concept of *Ubiquitous Research Preservation (URP)*. Additionally, we presented a framework composed of three models for data governance, management and sharing. We expect URP to make a positive impact on researchers' ability to reflect on past processes, to provide suitable training material and to improve the reproducibility of their work. Yet, we do not intend to oversimplify complex use cases. Preservation is a first step towards supporting those, but it is not the only requirement. For example, the decision to share resources in order to improve research reproducibility does not only depend on the effort to preserve data, but on various other factors, including competition and privacy policies.

Our research focuses on computational and data-driven scientific domains, as automated, machine-supported knowledge preservation promises to best map experimental processes and relevant resources. Yet, as all science became to varying degrees connected to computation, we expect URP to profit scientific domains beyond computational and data-driven science.

ACKNOWLEDGMENTS

This work is partially supported and funded by the following entities: the Amplify project (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 683008), and the HumanE AI Network under the European Union's Horizon 2020 ICT programme (grant agreement no. 952026) .

REFERENCES

[1] Anna Bánáti, Péter Kacsuk, and Miklós Kozlovszky. 2015. Four level provenance support to achieve portable reproducibility of scientific workflows. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 241–244.

[2] Christine L Borgman. 2007. *Scholarship in the digital age: information, infrastructure, and the internet.* MIT Press, Cambridge.

[3] Yi Chen and Gareth JF Jones. 2010. Augmenting human memory using personal lifelogs. In *Proceedings of the 1st augmented human international conference*. 1–9.

[4] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1143–1152.

[5] Sarah Clinch, Tilman Dingler, Kai Kunze, Passant ElAgroudy, and Tsutomu Terada. 2017. WAHM-4th workshop on ubiquitous technologies to augment the human mind: sharing experiences. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 662–665.

[6] Tilman Dingler, Passant EL Agroudy, Gerd Matheis, and Albrecht Schmidt. 2016. Reading-based screenshot summaries for supporting awareness of desktop activities. In *Proceedings of the 7th Augmented Human International Conference 2016*. 1–5.

[7] Tilman Dingler, Passant El Agroudy, Huy Viet Le, Albrecht Schmidt, Evangelos Niforatos, Agon Bexheti, and Marc Langheinrich. 2016. Multimedia memory cues for augmenting human memory. *IEEE MultiMedia* 23, 2 (2016), 4–11.

[8] David Donoho, Arian Maleki, Inam Rahman, Morteza Shahram, and Victoria Stodden. 2008. 15 years of reproducible research in computational harmonic analysis. *Technical report* (2008).

[9] Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, alt02.

[10] Passant Elagroudy, Sebastian S. Feger, and Albrecht Schmidt. 2022. A Model for Selecting Media Type of Memory Cues in Ubiquitous Prostheses. In *AHs 2022: Augmented Humans 2022, Hybrid Event / Munich, Germany, March 13 - 15, 2022*, Masaaki Mochimaru, Yuichi Kurita, Martin Schmitz, and Karola Marky (Eds.). ACM, 313–316. https://doi.org/10.1145/3519391.3524169

[11] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Albrecht Schmidt, and Paweł W. Woźniak. 2019. Designing for Reproducibility: A Qualitative Study of Challenges and Opportunities in High Energy Physics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'19* (2019). https://doi.org/10.1145/3290605.3300685

[12] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. 2019. Gamification in Science: A Study of Requirements in the Context of Reproducible Research. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'19* (2019). https://doi.org/10.1145/3290605.3300690

[13] Sebastian S Feger, Sünje Dallmeier-Tiessen, Paweł W Woźniak, and Albrecht Schmidt. 2019. The Role of HCI in Reproducible Science: Understanding, Supporting and Motivating Core Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[14] Sebastian S Feger, Paweł W Wozniak, Lars Lischke, and Albrecht Schmidt. 2020. 'Yes, I comply!' Motivations and Practices around Research Data Management and Reuse across Scientific Fields. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26.

[15] Angela Fessl, Aitor Apaolaza, Ann Gledson, Viktoria Pammer-Schindler, and Markel Vigo. 2019. "Mirror, mirror on my search...": Data-Driven Reflection and Experimentation with Search Behaviour. In *European Conference on Technology Enhanced Learning*. Springer, 83–97.

[16] Luke Haliburton and Albrecht Schmidt. 2020. Technologies for healthy work. *Interactions* 27, 3 (2020), 64–66.

[17] Morgan Harvey, Marc Langheinrich, and Geoff Ward. 2016. Remembering through lifelogging: A survey of human memory augmentation. *Pervasive and Mobile Computing* 27 (2016), 14–26. https://doi.org/10.1016/j.pmcj.2015.12.002

[18] Tony Hey, Stewart Tansley, and Kristin Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/

[19] Zaher Hinbarji, Moohamad Hinbarji, Rami Albatal, and Cathal Gurrin. 2016. Personal information manager to capture and re-access what we see on computers. In *Proceedings of the first Workshop on Lifelogging Tools and Applications*. 13–17.

[20] Buket Kip Kayabas and Mehmet Emin Mutlu. 2015. Record, evaluation and planning of knowledge work experiences on personal research environments via life logging system. *Procedia-Social and Behavioral Sciences* 174 (2015), 1129–1137.

[21] İlker Kayabaş and Mehmet Emin Mutlu. 2015. Obtainment and management of informal learning experiences among saved life experiences via a life logging system: an observation of a software developer. *Procedia-Social and Behavioral Sciences* 174 (2015), 1111–1116.

[22] Mary Beth Kery, Marissa Radensky, Mahima Arya, Bonnie E. John, and Brad A. Myers. 2018. The Story in the Notebook: Exploratory Data Science using a Literate Programming Tool. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'18* (2018), 1–11. https://doi.org/10.1145/3173574.3173748

[23] Hyowon Lee, Alan F Smeaton, Noel E O'connor, Gareth Jones, Michael Blighe, Daragh Byrne, Aiden Doherty, and Cathal Gurrin. 2008. Constructing a SenseCam visual diary as a media process. *Multimedia Systems* 14, 6 (2008), 341–349.

[24] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 557–566.

[25] Andrew Morin, Jennifer Urban, Paul D Adams, Ian Foster, Andrej Sali, David Baker, and Piotr Sliz. 2012. Shining light into black boxes. *Science* 336, 6078 (2012), 159–160.

[26] Gerard Oleksik, Natasa Milic-Frayling, and Rachel Jones. 2014. Study of electronic lab notebook design and practices that emerged in a collaborative scientific environment. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14* (2014), 120–133. https://doi.org/10.1145/2531602.2531709

[27] Victoria Stodden and Sheila Miguez. 2013. Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. (2013).

[28] Aurélien Tabard, Wendy E Mackay, and Evelyn Eastmond. 2008. From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 569–578.

[29] Kenji Tanaka, Hiroki Nakayama, Ryo Onuma, Hiroaki Kaminaga, Youzou Miyadera, and Shoichi Nakamura. 2019. Organization of Work Situations and the Involved Documents in Complicated Documenting via Intuitive Triggers. In *2019 IEEE Conference on Big Data and Analytics (ICBDA)*. IEEE, 1–5.

[30] Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. Data sharing by scientists: practices and perceptions. *PloS one* 6, 6 (2011), e21101.