Acta Scientific Microbiology (ISSN: 2581-3226)

Volume 8 Issue 1 January 2025

Research Article

# Science/Education Portraits X: Personal Archive of a Scientist

**Maurice HT Ling***

*HOHY PTE LTD, Singapore*

***Corresponding Author:** Maurice HT Ling, HOHY PTE LTD, Singapore.

**Received:** December 02, 2024

**Published:** December 21, 2024



## Abstract

In 2007, Sydney Brenner and Richard Roberts call for the archival of such non-published or non-publishable records, such as notes and drafts, as our scientific heritage. Using this as imperative, I reflect on my own scientific life and the records generated. These include research notebooks, research data, and auxiliary records (which includes emails, diaries, instant messages, among others). I concur and urge that every scientist should consider themselves as personal archivist to protect and preserve our collective heritage.

**Keywords:** Scientific Records; Research Notebooks; Research Data; Diaries; Instant Messages; Emails; Thoughts and Reflections

## Scientist as Archivist

Personal archives are defined by as Rob Fisher [1] as records created by individuals and corporate entities (including non-profit organizations) outside of the public sphere of governments, governmental agencies, and departments. This is consistent with Jennifer Douglas' view that the defining feature of a personal archive is that it was created by an individual (or individuals acting communally) rather than by an organization or corporate body [2]. Both definitions suggest that personal archive can be created individually or co-created communally – two of the five ways of archive creation [3]. The distinction between these two is blurred. While it can be conceived that personal notes and reflections, such as diary entries, are created individually; photographs and videos are usually co-created. In fact, it can be argued that even diary entries cannot be created in isolation. Catherine Hobbs [4] argues that personal archive has the nuances of the individual and his/her interpretation of the surroundings and values, as well as interactions within the society. As such, what is deemed as a personal archive is part of a communal archive, such as a family archive; which is instrumental in reconstructing history [5]. Indeed, Elizabeth van Heyningen [6] argues that personal diaries have significant historical value. For example, it is well-known that Anne Frank's diary provides a window into the common Jews hiding within Nazi occupied Holland during World War II [7–9], and is listed within Scott Christianson's "100 Documents That Changed the World: From the Magna Carta to Wikileaks" [10].

Integral to a scientist's work is records and data creation [11] in the form of research notebooks and computer files [12]. As maintaining research data is a necessity [13,14], a scientist has to take on the role of an archivist [15]. Scientific records go beyond the traditional research notebooks and may include less formal forms, such as emails and instant messaging [16]. This concurs with Sydney Brenner and Richard Roberts whom have called for the archival of such non-published or non-publishable records, such as notes and drafts, as our scientific heritage [17]. I personally finds this to be a worthy but highly unachievable goal at least for now at the moment as a study that I did with a student in 2014 to examine the databases published in 2013 found that 290 of the 379 (76.5%) database papers contained URLs to the database; worst of all, 18 of the 290 (6.62%) of the databases with URLs were either not accessible or intermittently accessible [18]. This suggests an attrition rate of 28.23%, as determined by accessibility, in less than 2 years post-publication.

Nevertheless, as a scientist and a history lover, I find myself instinctively gravitated towards the archivist role [19] and have allowed it to permeate throughout my life. Hence, this article is my

reflection on the interplay between the scientist and archivist in me, resulting in my current personal archive; as well as making sense of this archive.

## Research Notebook

Research notebook is the foundation stone of a scientist's work and identity, that any reflection of a scientist's archive must start with it. The term research notebook is synonymous with laboratory notebook; and in many instances, generally synonymous with research logbook or laboratory logbook. However, there are subtle and important differences. Semantically, laboratory notebook/logbook implies its use to record laboratory work while research notebook/logbook includes laboratory and non-laboratory work and thoughts. However, such distinction is usually blurred. In practice, one can generally consider all four forms (research notebook, research logbook, laboratory notebook, and laboratory logbook) to be synonymous. The importance of a laboratory notebook to a scientist cannot be understated [20,21] as it is a key component in scientific discovery [22] as it generally records actual work done in the laboratory setting. The U.S. Health and Human Services Office of Research Integrity summing up its importance as (URL 1):

- To establish good work practices
- To teach the people in your lab
- To meet contractual requirements
- To avoid fraud
- To defend patents
- To allow work to be reproduced by others
- To facilitate preparation of formal reports, presentations and papers
- To validate your research
- To serve as a source for assigning credit to lab members

Historically, research/laboratory notebooks have played critical roles in the establishing milestones in science [23]; resolving controversies, such as Louis Pasteur's anthrax vaccine [24]; re-analysis of historical data [25], and establishing inventorship (such as in the case of University of Pittsburgh v. Hedrick, 573 F.3d 1290 (Fed.Cir.2009)). Research/laboratory notebooks can be an important piece of puzzle to minimize scientific fraud [26,27], such as data fabrication and data falsification. Fanelli [28] performed a meta-analysis on 21 published surveys between 1987 and 2008, amounting to 11647 respondents, and found that 14.12% of respondents have personal knowledge of a colleague who fabricated or falsified research data, or who altered or modified research data. I performed a follow up meta-analysis on 10 published surveys between 2009 to 2018 suggested 17.7% of the respondents have knowledge of fellow scientist's acts of data fabrication or data falsification [29]. As a result, many guidelines [30,31] have been written with regards to proper notebook keeping.

As a scientist and research mentor, I often must guide new research students on proper research/laboratory notebook keeping. Over time, I summarize several guidelines for myself and my students:

- Number every page, and no page(s) should be torn out of the notebook.
- Write all entries in ink.
- Every entry must be (a) dated, and (b) complete with objectives, procedures, and results. Essentially, you or your friend should be able to replicate the experiment/work from your entry.
- Do not write into the margins. This is to leave space for binding.
- Do not leave any empty pages.
- If you make a mistake, don't obliterate it! Strike out the error while leaving the error readable.
- Never use correction tape or fluid – this is equivalent to scientific crime.
- Write down the file name(s) of any computer files used or generated, including any written computer codes. The computer files should be readily identifiable from notebook entries. As much as possible, these computer files should be version controlled and printed.
- Printouts must be firmly attached to the page.
- Have your notebook notarized routinely.

To cap it off, I add that "your research or laboratory notebook/logbook is your primary record of all your work, readings, and interpretations of your work. The term logbook comes from explorer's or mariner's log where each measurement and direction are kept throughout the journey – for complete replotting of the route, if necessary. Hence, your research or laboratory notebook/logbook must be maintained as such. In event of litigation, your research or laboratory notebook/logbook is likely the first piece of evidence to be called for in the court of law".

Although the ownership of research notebooks can belong to the principal investigator, institution, or research sponsor; the owner may allow the researcher to make a copy when he/she leaves the laboratory or research group. I have heard instances of mentor-mentee conflicts surrounding this issue where the mentee wants a photocopy of his/her research notebooks but disallowed by the mentor due to many possible reasons; and quite often, this conflict is never fully resolved resulting in animosity. Nevertheless, this suggests that a scientist's research notebooks comprise of two series – the scientist's own research notebooks, and research notebooks of the scientist's students or mentees. This further implies that a scientist's own research notebooks (the former series) is likely to have its roots in or as research notebooks of his/her research mentor or adviser (the latter series) – a parallel to academic genealogy [32,33]. The transition between the latter to the former is usually parallel running or direct changeover, rather than phase implementation. In parallel running, the original research notebooks may be kept by the advisor as the advisor's latter series with a duplicated or imaged version as the start of mentee's own research notebook series. In direct changeover, an early career scientist or graduate student may stop contributing to his/her advisor's latter series entirely while keeping his/her research notebooks.

I have the privilege of a direct changeover after my honours year when I move from molecular biology and cell biology research into bioinformatics, with the first entry dated 7th May 2004 (Figure 1). This entry recorded the discussion I had with Dr. Christophe Lefevre, who was with Victorian Bioinformatics Consortium [34] (https://github.com/Victorian-Bioinformatics-Consortium) and would be my PhD co-supervisor eventually, on the potential directions that I could take for my doctoral studies which would officially begin in July 2004.

Looking back, I found a subtle shift in research notebooks moving from wet-laboratory research to dry-laboratory research. When I was doing wet-laboratory research during my honours year; my supervisor, Dr Kevin Nicholas; is insistent in research notebooks. In fact, we must use the official issued laboratory notebooks as Kevin refuses to read or discuss about anything not written in there. However, Kevin is much less insistent on laboratory notebooks when I started my doctoral candidature in bioinformatics with him. As a result, I am able to start my own research notebook series. I think one of the fundamental features of bioinformatics work is the presence of source codes written for the analysis of data. Source codes are both human and machine readable [35]; which meant that if the data and source codes for process and analysis are present, the work is documented. This is supported by recent studies in automated documentation generation from source codes [36,37]. Hence, the presence of source code files for data analysis can be deemed as "according to manufacturer instructions" in many publications. I had also archived the codes from my doctoral studies (URL 2) in Software Heritage [35], which aims to "collect, preserve, and share all software that is publicly available in source code form".

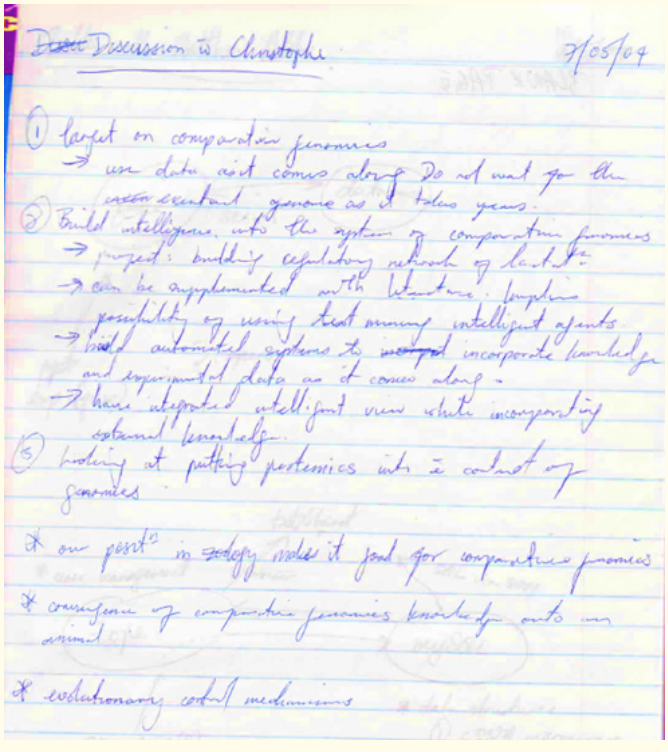



**Figure 1:** The First Page of My Research Notebook. This entry is dated May 07, 2004; containing the minutes of discussion with Dr. Christophe Lefevre, my eventual PhD co-supervisor, on my potential the potential directions for my PhD.

Since May 07, 2004 (Figure 1), I have accumulated 3800 A4 pages of research notebooks for myself (Figure 2A) and gathered nearly 4000 A4 pages of research notebooks from my students (Figure 2B). On a personal standpoint, it has been known that research notebooks are tools to consolidate learning [38]. From the standpoint of archival, there are intricate details embedded within the research notebooks about the researcher's working style [39] and thought processes [40]. Recording of research data is not a

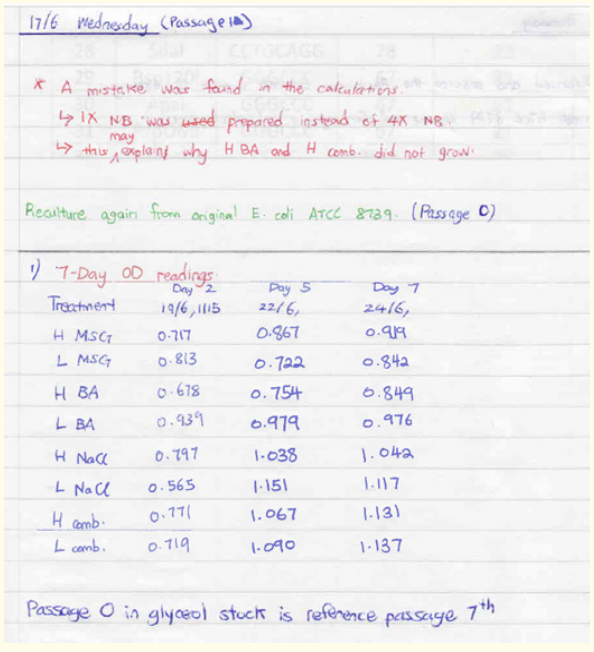
haphazard process but a well-considered process. For example, the same data can be recorded as a list or in a table, which has drastic differences in subsequent analyses. Similarly, using different coloured ink to categorize thought processes can demonstrate high-level thinking (Figure 3A). Creativity and artistic talent can also be demonstrated for bioinformatics work (Figure 3B). Personally, I find that research notebooks is an invaluable asset attesting to the work done.

A


B


**Figure 2:** Research Notebooks. Panel A shows 14 volumes of my research notebooks, spanning 3800 pages. Volume 12 is still in pre-bounded state. Panel B shows 9 volumes of collated research notebooks of my students, spanning 3951 pages. These research notebooks are A4 sized.

A




B

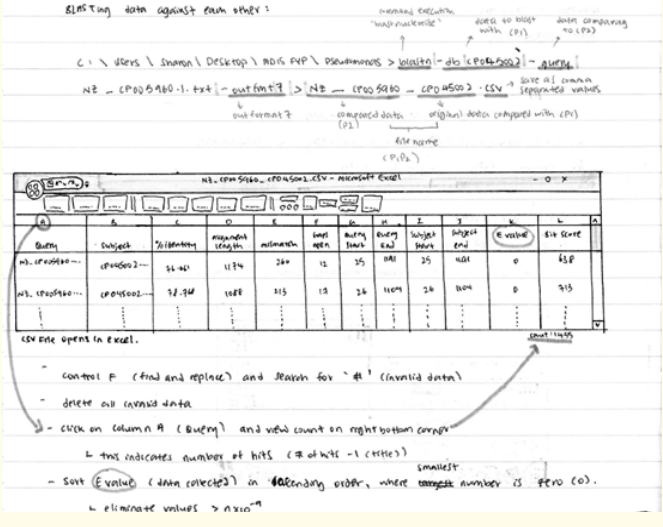



**Figure 3:** Examples of Research Notes from Students. Panel A shows the use of different colours to categorize annotations – red for mistakes and important headers, blue for data, and green for information. Panel B is an example of creative and artistic talent while doing bioinformatics work.

## Research data

Extending from research notebooks is actual research data – data that eventuates or instrumental in a publication. Although a good proportion of journals encourage data availability and sharing by 2020 [41], it is hardly sufficient [42] as a study by Gabelica., *et al*. [43] found that they can only successfully obtain research data from 6.8% of the published papers with data availability statement. More importantly, 85.8% (1538 of 1792) did not respond to data sharing request despite indicating in their publications that they are willing to share their data. Data availability is often the first step towards reproducibility [44].

A study by Federer [45] suggests that availability of URLs or DOIs to retrieve data may be more trustworthy that a statement of willingness to share data, which I am supportive of. As such, I aim to release as much data as feasibly possible at my own website (URL 3) for each publication. Moreover, I try my best to include the necessary data files, such as downloaded sequence data, in my data release (Figure 4). While this may seem unnecessary at that point in time since I can always put-up URLs to download the pre-requisite files, I realize that this to be an appropriate move when I attempt to retrieve data from dead URLs given the published papers of others – my email enquiries to the authors are not replied; thus, rendering the publication unusable. At the same time, I find that by making research data publicly available pushes me to archive my data in a more usable and neater manner.


**Figure 4:** Sample of Research Data Release. MapMan-Tobacco.rar is the main research data file release. However, I choose to release Tobacco Gene Index release 1 to 7, in case they cannot be found later.

## Auxiliary records

Circumnavigating research notebooks and research data is a vast space of auxiliary records, which includes emails and instant messaging [16], notes and drafts [17]. This almost covers a scientist's entire written or typed text, which are worth preserving [17] and I concur. While preparing this manuscript, I flipped through pages of my diaries (Figure 5) and schedules (Figure 6) and found both frustrations of failed experiments, thought processes, as well as the elation of successes in my career – none of which can be published in any scholarly form. Yet, within these pages contains the vivid experiences as they happen [46], the links between projects and the thought processes that glues them together; as Lotte Mulligan found in the case with Robert Hooke's diaries [47]. Within my instant messaging logs are discussions with my colleagues, friends, and ex-students about science and how we approach science, how our scientists' work affects our thoughts, our understanding, similarities, differences, and notably, reflections. All of which can be considered duoethnographic in nature [48,49]. For these reasons, I feel that there is value in its preservation and hopefully, contribution to our collective scientific heritage.


**Figure 5:** Thirty-Six Volumes of Diaries/Journals. The diaries/journals are A5 sized, spanning 10985 pages, from 17 March 2008 to 23 August 2024 – 6004 days; which equates to 16 years, 5 months, and 7 days.

**A**


**B**

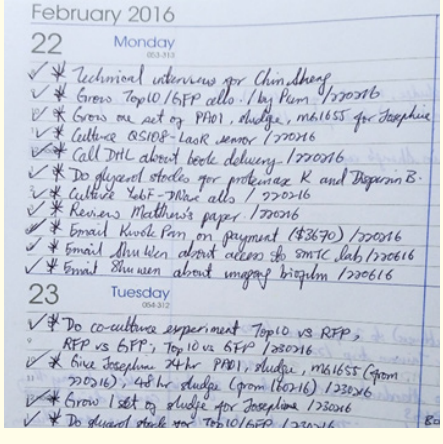



**Figure 6:** Nine volumes of Schedules. Panel A shows 9 volumes of annual schedules in the form of management diaries from 2016 to 2024. Panel B shows a page from 22/23 February 2016.

## Concluding Remarks

In the National Palace Museum of Taiwan lies an artifact known as "Lord Mao's cauldron", created at around 805 BC, and is historically important due to its inscriptions of 500 Chinese characters – the longest ancient Chinese bronze inscriptions known today (URL 4). The work of a scientist consists much more than published papers. Other unpublished or unpublishable materials; such as, research notebooks, research data, thoughts, formal and informal communications; are also part of our collective scientific heritage and worth preserving. Science is a large collaborative human endeavour [50,51]; hence, every scientist should consider themselves as personal archivist.

## Bibliography

1. Fisher R. "In Search of a Theory of Private Archives: The Foundational Writings of Jenkinson and Schellenberg Revisited". *Archivaria* 67 (2009): 1-24.

2. Douglas J. "Getting Personal: Personal Archives in Archival Programs and Curricula. *Education for Information* 33.2 (2017): 89-105.

3. Douglas J. "A Call to Rethink Archival Creation: Exploring Types of Creation in Personal Archives". *Archival Science* 18.1 (2018): 29-49.

4. Hobbs C. "The Character of Personal Archives: Reflections on the Value of Records of Individuals". *Archivaria* 52 (2001): 126-135.

5. Schwartzenberg S. "The Personal Archive as Historical Record". *Visual Studies* 20.1 (2005): 70-82.

6. Van Heyningen E. "The Diary as Historical Source: A Response". *Historia* 38.1 (1993): 14-25.

7. Doneson JE. "The American History of Anne Frank's Diary. *Holocaust and Genocide Studies* 2.1 (1987): 149-160.

8. Hasian MA. "Anne Frank, Bergen-Belsen, and the Polysemic Nature of Holocaust Memories". *Rhetoric and Public Affairs* 4.3 (2001): 349-374.

9. Foray JL. "The Nation Behind the Diary: Anne Frank and the Holocaust of the Dutch Jews". *The History Teacher* 44.3 (2011): 329-352.

10. Christianson S. "100 Documents That Changed the World: From the Magna Carta to Wikileaks (Batsford Ltd, Great Britain)" (2015).

11. Matias JN., *et al*. "The Upworthy Research Archive, A Time Series of 32,487 Experiments in U.S. Media". *Scientific Data* 8.1 (2021): 195.

12. Kanza S and Knight NJ. "Behind Every Great Research Project is Great Data Management". *BMC Research Notes* 15.1 (2022): 20.

13. Syn SY and Kim S. "Characterizing the Research Data Management Practices of NIH Biomedical Researchers Indicates the Need for Better Support at Laboratory Level". *Health Information and Libraries Journal* 39.4 (2022): 347-356.

14. Birkbeck G., *et al*. "Challenges in Research Data management Practices: A Literature Analysis". *Journal of Decision Systems* 31 (2022): 153-167.

15. Borgerud C., *et al*. "Open Research Data, An Archival Challenge?" *Archival Science* 20.3 (2020): 279-302.

16. Laloë A-F. "Archives Of And For Science: Archives for Molecular Biology Preserve the Heritage of Science Beyond the Published Record for Future Scholars". *EMBO Reports* 18.8 (2017): 1273-1278.

17. Brenner S and Roberts RJ. "Save Your Notes, Drafts and Printouts: Today's Work is Tomorrow's History". *Nature* 446.7137 (2007): 725-725.

18. Koh YZ and Ling M. "Catalog of Biological and Biomedical Databases Published in 2013". *iConcept Journal of Computational and Mathematical Biology* 3 (2014): 3.

19. Cook J. "The Archivist: Link Between Scientist And Historian". *The American Archivist* 34.4 (1971): 377-381.

20. Hills S and Light CJ. "Training for Responsible and Ethical Management of Lab Notebooks in a Course-Based Undergraduate Research Experience". *Journal of Microbiology and Biology Education* 23.2 (2022): e00024-22.

21. Hopen A. "Proving Inventorship: The Importance of the Inventor/Laboratory Notebook". *Retina (Philadelphia, Pa)* 25.8 (2005): S99.

22. Nussbeck SY., *et al*. "The Laboratory Notebook in the 21st Century: The Electronic Laboratory Notebook Would Enhance Good Scientific Practice and Increase Research Productivity". *EMBO Reports* 15.6 (2014): 631-634.

23. Buchwald JZ. "Reworking the Bench: Research Notebooks in the History of Science". eds Holmes FL, Renn J, Rheinberger H-J (Springer Netherlands, Dordrecht), 7 (2003).

24. Smith KA. "Wanted, An Anthrax Vaccine: Dead or Alive?" *Medical Immunology* 4.1 (2005): 5.

25. Roelofs A. "One Hundred Fifty Years After Donders: Insights from Unpublished data, A Replication, and Modeling of His Reaction Times". *Acta Psychologica* 191 (2018): 228-233.

26. Angelides K and Pianelli JV. "It Will Take More Than Notebooks to Stop Fraud". *Nature* 396.6710 (1998): 404-404.

27. van Steensel B. "Scientific Honesty and Publicly Shared Lab Notebooks: Sharing Lab Notebooks Along with Publication Would Increase Transparency and Help to Improve Honesty When Reporting Results". *EMBO Reports* 19.10 (2018): e46866.

28. Fanelli D. "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data". *PloS One* 4.5 (2009): e5738.

29. Ling MH. "Science/Education Portraits III: Perceived Prevalence of Data Fabrication and/or Falsification in Research". *Advances in Biotechnology and Microbiology* 11.5 (2020): 555824.

30. Schnell S. "Ten Simple Rules for a Computational Biologist's Laboratory Notebook". *PLoS Computational Biology* 11.9 (2015): e1004385.

31. Nickla J and Boehm M. "Proper Laboratory Notebook Practices: Protecting Your Intellectual Property". *Journal of Neuroimmune Pharmacology* 6.1 (2011): 4-9.

32. Jackson A. "A Labor of Love: The Mathematics Genealogy Project". *Notices of the American Mathematical Society* 54.8 (2007): 1002-1003.

33. Bennett AF. "The Academic Genealogy of George A. Bartholomew". *Integrative and Comparative Biology* 45.2 (2005): 231-233.

34. Seemann T. "Victorian Bioinformatics Consortium: Providing Essential Support for Victorian Science". *Asia-Pacific Biotech News* 10.24 (2006): 1400-1402.

35. Abramatic J-F., *et al*. "Building the Universal Archive of Source Code". *Communications of the ACM* 61.10 (2018): 29-31.

36. Sulír M and Porubän J. "Source Code Documentation Generation Using Program Execution". *Information* 8.4 (2017): 148.

37. Arthur MP. "Automatic Source Code Documentation using Code Summarization Technique of NLP". *Procedia Computer Science* 171 (2020): 2522-2531.

38. Glassman JA and Worsham DM. "Digital Research Notebook: A Simple Tool for Reflective Learning". *Reference Services Review* 45.2 (2017): 179-200.

39. Todes DP. "From Lone Investigator to Laboratory Chief: Ivan Pavlov's Research Notebooks as a Reflection of His Managerial and Interpretive Style". Reworking the Bench: Research Notebooks in the History of Science, Archimedes., eds Holmes FL, Renn J, Rheinberger H-J (Springer, Dordrecht) (2003): 203-220.

40. Buchwald JZ. "The Scholar's Seeing Eye". Reworking the Bench: Research Notebooks in the History of Science, Archimedes". eds Holmes FL, Renn J, Rheinberger H-J (Springer, Dordrecht) 7 (2003): 309-325.

41. Kim J., *et al*. "Data Sharing Policies of Journals in Life, Health, and Physical Sciences Indexed in Journal Citation Reports". *PeerJ* 8 (2020): e9924.

42. Bradley SH. "Data Availability Statements: A Little Credit, But Not Much". *BMJ* 380 (2023): o3047.

43. Gabelica M., *et al*. "Many Researchers Were Not Compliant With Their Published Data Sharing Statement: A Mixed-Methods Study". *Journal of Clinical Epidemiology* 150 (2022): 33-41.

44. Sanchez Reyes LL., *et al*. "Approachable Case Studies Support Learning and Reproducibility in Data Science: An Example from Evolutionary Biology". *Journal of Statistics and Data Science Education* 30.3 (2022): 304-310.

45. Federer LM. "Long-Term Availability of Data Associated with Articles in PLOS ONE". *PLoS One* 17.8 (2022): e0272845.

46. Bolger N., *et al*. "Diary Methods: Capturing Life as it is Lived". *Annual Review of Psychology* 54.1 (2003): 579-616.

47. Mulligan L. "Self-Scrutiny and the Study of Nature: Robert Hooke's Diary as Natural History". *Journal of British Studies* 35.3 (1996): 311-342.

48. Sung R-J., *et al*. "Constructing Biology Education Research Identities: A Duoethnography". *Frontiers in Education* 8 (2023): 1134040.

49. Shin AW., *et al*. "Science/Education Portraits VIII: Duoethnography of First-Generation Bioscience Undergraduates in a Private Education Institute in Singapore". *Acta Scientific Microbiology* 6.6 (2023): 24-35.

50. Ellemers N. "Science as Collaborative Knowledge Generation". *British Journal of Social Psychology* 60.1 (2021): 1-28.

51. Paige K and Hardy G. "Science as Human Endeavour, Critical Pedagogy and Practitioner Inquiry: Three Early Career Cases". *International Journal of Science and Mathematics Education* 17.4 (2019): 679-699.

## List of URLs

**URL 1:** https://ori.hhs.gov/education/products/wsu/data_lab.html

**URL 2:** https://archive.softwareheritage.org/browse/origin/directory/? origin_url=https://svn.code.sf.net/p/muscorian/code/

**URL 3:** https://github.com/mauriceling/mauriceling.github.io/wiki/Data-Set-Downloads

**URL 4:** https://en.wikipedia.org/wiki/Mao_Gong_ding