Introducing the MusicLab Copenhagen Dataset

Simon Høffding[1,2,3] (iD)**, Rebecca Josefine Five Bergstrøm**[4]**,
Laura Bishop**[1,2] (iD)**, Pedro Lucas**[1,5]**, Kayla Burnim**[1,2]**,
Carlos Cancino-Chacón**[6]**, Alena Clim**[1,2]**, Mathew Good**[4] (iD)**,
Niels Chr. Hansen**[7,8,9] (iD)**, Eirik Slinning Karlsen**[1,2]**,
Live Håndlykken Kvale**[4] (iD)**, Bruno Laeng**[1,10]**, Olivier Lartillot**[1,2]**,
Eigil Lippert**[11] (iD)**, Remy Martin**[1,2]**, Nanette Nielsen**[1,2] (iD)**,
Rahul Omprakash**[1,10]**, Thea Sørli Paulsrud**[1]**, Fernando Rosas**[12]**,
Dana Swarbrick**[1,2] (iD)**, Solveig Sørbø**[4]**, Finn Upham**[1,2] (iD)**,
Atilla Vrasdonk**[1,10] (iD)**, Jonna Vuoskoski**[1,2,10] (iD)**,
Sebastian Wallot**[13] (iD)**, Wenbo Yi**[1,2] (iD)**, Anne Danielsen**[1,2] (iD)
and Alexander Refsum Jensenius**[1,2] (iD)

Abstract

MusicLab Copenhagen was a unique research concert featuring the world-renowned Danish String Quartet in a naturalistic setting. The audience was split between one group physically located in the hall, another group listening to a radio broadcast, and a third group watching a live stream. Qualitative and quantitative data were captured from both musicians and audiences, resulting in a comprehensive dataset that can be used to address many research questions. This document introduces the dataset, explains its structure, and reflects on the related data collection, storing, publishing, and archiving processes.

Keywords

Citizen science, concert research, data paper, FAIR data, MusicLab Copenhagen, open research

[1] RITMO Centre for Interdisciplinary Studies in Rhythm, Time, and Motion, University of Oslo, Oslo, Norway

[2] Department of Musicology, University of Oslo, Oslo, Norway

[3] "Movement, Culture, and Society", Department of Sports Science and Biomechanics, University of Southern Denmark, Odense, Denmark

[4] University of Oslo Library, Oslo, Norway

[5] Department of Informatics, University of Oslo, Oslo, Norway

[6] Institute of Computational Perception, Johannes Kepler University Linz, Linz, Austria

[7] Centre of Excellence in Music, Mind, Body, & Brain, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland

[8] Royal Academy of Music Aarhus, Aarhus, Denmark

[9] Aarhus Institute of Advanced Studies, Aarhus University, Aarhus C, Denmark

[10] Department of Psychology, University of Oslo, Oslo, Norway

[11] Technical University of Denmark, Kgs. Lyngby, Denmark

[12] Department of Brain Sciences, Imperial College London, London, UK

[13] Institute of Psychology, Leuphana University of Lüneburg, Lüneburg, Germany

Corresponding author:

Simon Høffding, "Movement, Culture, and Society", Department of Sports Science and Biomechanics, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark.

Email: simonhof@imv.uio.no

Introduction

This paper introduces the large dataset of a unique musical event organized in Copenhagen, Denmark, on October 26, 2021. This event, MusicLab Copenhagen, was based on a collaboration between researchers from the University of Oslo and the world-renowned Danish String Quartet (DSQ). The quartet agreed to turn one of their public concerts into a research concert. It was initially planned for 2020 but was postponed several times due to the COVID-19 pandemic. While the delay was unfortunate, it also allowed the researchers to prepare for an even larger data collection than initially planned. Approximately 20 researchers, engineers, and assistants were involved in planning and executing the research component of the event. Additional DSQ staff, a concert production team, a documentary film team, and a Danish Broadcasting Corporation (DR) production team were also involved. Such a collaborative effort is unusual in music research, and it has inspired a new way of working among the team members (Danielsen et al., 2023).

For an overview of the objectives, research questions, and hypotheses behind MusicLab Copenhagen, see the editorial to the *Music & Science* special collection on "MusicLab Copenhagen: A research concert with the Danish String Quartet" (Høffding et al., 2024). The special collection also contains several papers describing and analyzing various parts of the dataset. These are not exhaustive; the dataset is so rich that our team can never analyze everything comprehensively. We are eager to share it openly to allow others to explore the material.

The MusicLab Copenhagen Dataset is shared on Open Science Framework (OSF)[1] with a permissive license (CC-By Attribution 4.0 International). The ambition was to share everything openly, but, as we will discuss later, some parts cannot be shared due to privacy and/or copyright limitations. Instead, we have made the dataset as "open as possible but as closed as necessary," following the FAIR principles to make the data findable, accessible, interoperable, and reusable (Wilkinson et al., 2016). Sharing is done in the spirit of Open Research. We prefer the term "Open Research" over "Open Science" to underline that we work between art and science and with many theoretical and methodological approaches.

The MusicLab Copenhagen Dataset includes a detailed data management plan[2] containing specific data capture and preprocessing information. It also provides contextual information to help understand more about the data, as a way to capture both *metadata* and *paradata*. Huvila et al. (2024) argue that contextual information, in the form of paradata, is crucial for more complex humanities-oriented data to be meaningful. For copyright reasons, we cannot share the complete video recording of the event on OSF, but the broadcast stream recording is available on YouTube.[3] Additionally, a short documentary film, *The Sound of Consciousness*,[4] shows snippets from the performance and interviews with involved researchers and musicians.

Guidelines for data papers are currently being developed in many disciplines (Walters, 2020), and there is also a growing interest in the music research community (Moss & Neuwirth, 2021). This data paper shows how complex, music-related datasets can be developed and shared, and we hope to inspire others to do the same.

Event Overview and Preparation

The MusicLab Copenhagen dataset consists of two parts: The Morning Experiment and The Evening Concert. Both were carried out on the same day and in the same hall, Musikhuset, an 18th-century mansion in central Copenhagen. Only the experimenters were in the hall during The Morning Experiment, while a full audience was present during The Evening Concert. The technical setup was rigged up early in the morning and was taken down late at night on October 26, 2021. Due to time constraints, the setup was meticulously planned via excellent organizational teamwork among the almost 40 people involved in the production. This included numerous online meetings between the key teams (research, DSQ, and production) and several workshops among the researchers to determine the data capture details based on research questions. Also, the team organized an event (MusicTestLab[5]) in Oslo in October 2020 to test the equipment with a student string quartet, the Borealis String Quartet, a student ensemble from the Norwegian Academy of Music in Oslo. This event was streamed live, with commentary, and many key researchers discussed methodological challenges. As such, much was learned and documented even before we came to Copenhagen.

The Morning Experiment

The Morning Experiment replicated a prior session in the fourMs lab at the University of Oslo with the Borealis String Quartet. They performed an excerpt from the first movement of Haydn's String Quartet in B-flat major, Op. 76, No. 4 under different conditions, stressing their communication and sense of togetherness (Bishop et al., 2021).

During their take on the task, The Danish String Quartet musicians wore heart rate sensors (Delsys Trigno EKG) under lycra suits with reflective markers tracked by an infrared motion capture system (Optitrack Flex 13). They also wore eye-tracking glasses (PupilLabs Core). Multiple audio recorders and video cameras captured the stage from different angles.

The experimental conditions are summarized in Figure 1 (see example of a condition in Figure 2). They entailed manipulating what the musicians saw during each performance of the piece: (1) playing back-to-back so that they were unable to see each other ("Blind"), (2) looking only at the score ("Score-directed"), (3) playing normally ("Normal-rehearsal"), (4) putting a screen in front of the first violin ("Violin-isolated"), (5) playing normally again as a control ("Replication-rehearsal"), and (7) playing as if it were a concert ("Concert"). In addition, the DSQ was asked to (6) sight-read an unknown piece ("Sight-reading"), Rued Langgaard's String Quartet no. 5, 2nd movement, to see how they acted in a more stressful situation.
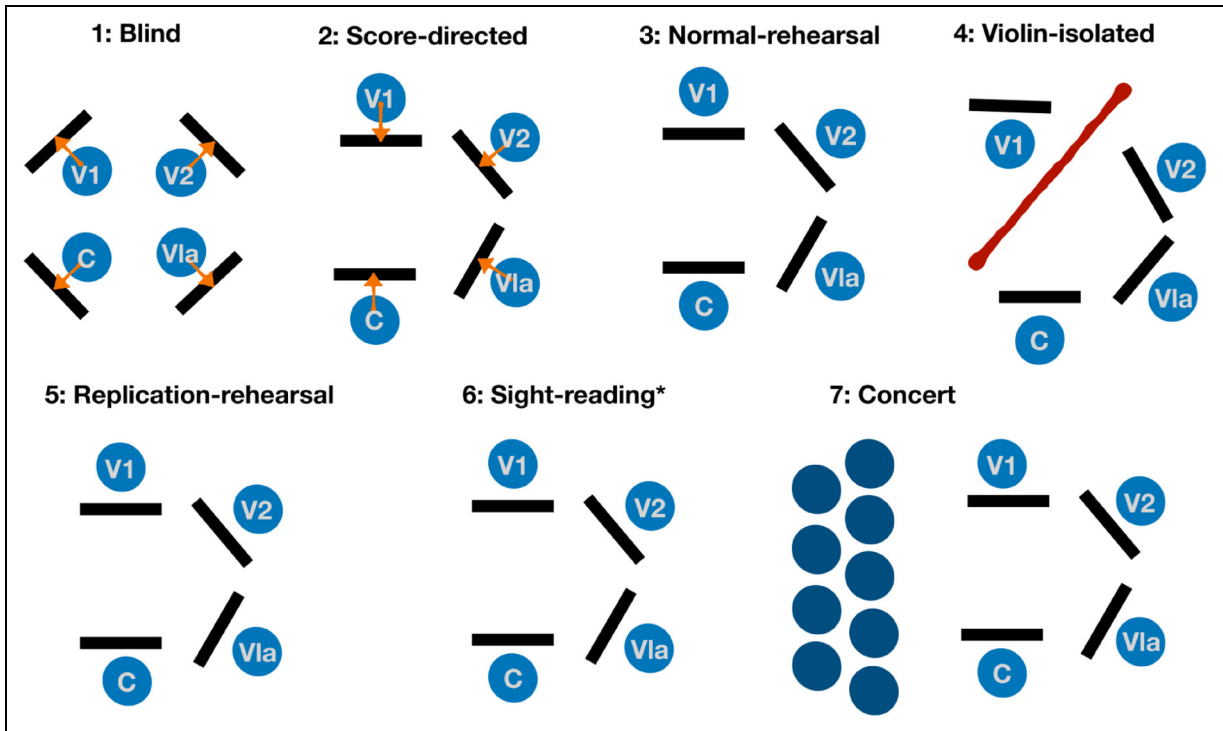
Figure 1. Performance conditions used in the replication study in The Morning Experiment included in the MusicLab Copenhagen Dataset (the figure is a reproduction from Bishop et al., 2023).

Figure 2. The DSQ playing with a screen between the first violinist and the rest of the ensemble during The morning experiment.

There was no regular audience during The Morning Experiment, but around 10 researchers sat in the hall, clapping after each condition. Several researchers explicitly stated how well they thought the DSQ played and how fortunate they felt having the opportunity to attend a "private concert" with them.

The Evening Concert

The second part of the dataset is from The Evening Concert. It was announced as a "research concert," and audience members were encouraged to participate in the research by downloading the MusicLab App[6] when purchasing
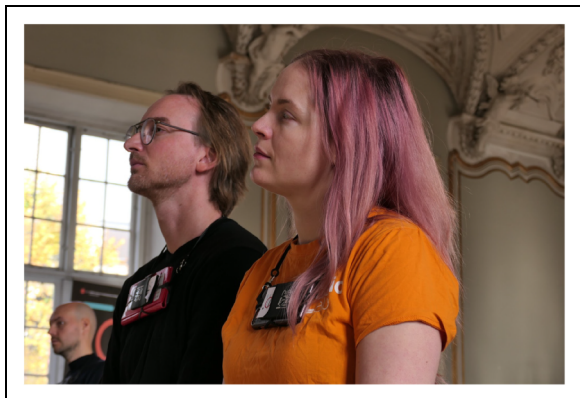
Figure 3. Motion data from audience members was captured from mobile phones hanging around people's necks running the custom-built MusicLab app.

tickets. This custom-built research app for Android and iOS phones allows for answering questions and capturing sensor information from mobile phones. This was a voluntary option for the audience, and those who agreed to participate received assistance to mount the phone in a custom-designed phone strap (Figure 3). Those without (compatible) smartphones could wear an accelerometer with a data logger (Axivity AX3) hanging around their neck.

All in-house audience members were handed a pamphlet with questionnaires to complete. We opted for a paper-based solution since their mobile phones were already in use for data capture.

Audience members who listened to the radio broadcast or watched the live stream on YouTube or Facebook were encouraged to hang their mobile phones around their necks with a strap, similar to the positioning of the in-house audience members. They used the MusicLab app to answer the same questions the in-house audience had on paper.

The concert started at 19:30 and consisted of two parts, followed by after-concert activities:

- Introduction speech and synchronization of recording devices
- Part 1: (~54 min)
  - Ludwig van Beethoven, String Quartet No. 16 op. 135
  - Alfred Schnittke, String Quartet No. 3
- Break (~30 min)
- Part 2: (~46 min)
  - Johann Sebastian Bach, Kunst der Fuge, Contrapunctus XIV
  - Folk music pieces (1) Mable Kelly / Planxty Kelly / Carolan's Quarrel with the Landlady (2) Stedelil (3) Halling etter Haltegutten (4) Unst Boat Song (5) Lovely Joan (6) Halling
- After-concert activities (~60 min)
  - Discussions with researchers and musicians
  - Data Jockeying

At the beginning of the concert, the principal investigator, Simon Høffding, spent 10 min introducing the background and purpose of the event without revealing too much about the data to be captured (cf. Danielsen et al., 2023). The introduction ended with group synchronization based on a finger-tapping cue (Upham, 2023). This synchronization procedure was repeated after the break. Before the concert and between each piece in the program, short "questionnaire-filling" sessions were inserted.

Part 1 of the concert was as close to a conventional concert as possible. The DSQ performed with heart rate sensors (Delsys EKG) under their shirts, which were invisible to the audience. Cameras and microphones were placed discreetly to disturb as little as possible.

The break was relatively long (~30 min), giving the audience time to work on their questionnaires and have an actual break. We also needed an extended break to prepare for the more extensive data collection protocols employed in Part 2.

The setup in Part 2 looked more "sciency" because musicians wore motion capture suits and eye-tracking glasses (Figure 4). During the J.S. Bach fugue, the musical structure was visualized on a screen next to the stage.

Immediately following the concert, the researchers and musicians mingled with the audience for about an hour in the foyer. In one of the rooms, researchers presented preliminary analysis as part of a "data jockeying" session.

Data Lifecycle

Most data was first collected locally on various capture devices or computers. Immediately following the event, all data was transferred to a university-managed external hard drive, where it was saved in a "raw" data folder divided into one subfolder per device according to the "good enough" approach suggested by Jensenius (2021). After transfer and preprocessing were confirmed, data was deleted from all capture devices. Next, data was preprocessed, converted with appropriate software, and saved in a "cooked" folder. All data from the research drive was uploaded to the secure UiO server once we could access a fast internet connection. The folders with raw and preprocessed data were made read-only on the server to avoid erroneous overwrites during further processing. Researchers copied data to their own folders on the server for analysis. The curated dataset on OSF consists of a subset of data from the various folders.

Due to the experiment's complex nature and the different local clocks on devices, synchronizing different data streams was complicated and required much manual work. Procedures and scripts have been shared,[7] reducing the post-processing needed for future data collection.

Whenever possible, time is synchronized to "concert time" in seconds from the first tapping synchronization cue performed during the concert. These tapping cues were identifiable in the motion measurements of the performers and audience. They can also be heard in the audio and seen in the video recordings of the event.

Figure 4. The Danish String Quartet performing with heart rate measurement devices, motion capture suits, and eye-tracking glasses during MusicLab Copenhagen on October 26, 2021.

The Dataset

The dataset on OSF is organized into four folders: Documentation, Private data, Published data, and Dissemination. We will discuss some of the content in the following sections.

Documentation

This Documentation folder contains the notification and approval documents from the Norwegian Centre for Research Data (NSD), reference number 915228. It also includes relevant information, including program notes and advertisements, that may help clarify the event and its organization.

The data management plan (DMP) is the most critical document in the Documentation folder. This is a living document that many people have contributed to writing. We used a shared online document with a customized template and encouraged everyone involved to update it regularly throughout the project. Over the years, we have experimented with different approaches to writing DMPs. There are benefits to having one person in charge of the document. However, asking everyone to contribute helped capture as many facets of the project as possible for such a large-scale, collaborative project. The risk is that everyone forgets about writing, and no one feels responsible for the document. Frequent reminders and peer pressure helped mitigate the problem.

A DMP should be a living document; as such, it will continue to be updated as the data is processed and analyzed. Furthermore, as more people use the data, we are becoming aware of shortcomings in the DMP that need to be addressed. The aim is to continue developing the DMP on our server to ensure adequate transparency and usefulness of the data. When changes are made, a PDF version of the DMP is exported and uploaded to OSF.

Private Data

This dataset folder has two purposes. One is a "staging area" to test individual data sections before publication. Once technical bugs are worked through, this section is replicated in the public folder. The second purpose is to store files that cannot be shared openly due to privacy and/or copyright restrictions. UiO researchers can access data on our university server. Other researchers granted access to "closed" data can access it in the Private folder on OSF. We are still working on obtaining permission to share some of the copyright-protected material, so hopefully, more data can be moved to the public folder over time.

Published Data

This folder is used to share anonymized audience data, musician sensor data, and recordings of the pieces that can be shared openly. We have tried to label folders and files to be self-explanatory, with additional information available in the data management plan, the wiki section, and readme files accompanying specific data.

Dissemination

This folder contains various output types, including scientific presentations, documentation of outreach activities, the documentary mentioned above, and the very articles included in this special collection. It primarily includes our material, as we will most likely not use this folder to share other people's use of the data.

Discussion

MusicLab Copenhagen was a massive investment for our research team, yet it has proven invaluable for building methodological expertise in the complex capture of

experts in a real-world context. The knowledge gained has enabled us to scale up to capturing full-scale symphony orchestras in subsequent years. However, as we have learned from MusicLab Copenhagen, capturing the data is only the beginning of the process. Pre-processing all the different data and media files is more time-consuming than one can imagine. Then comes the stage of synchronizing all the various file types to be compared in later analyses. Owing to the number of people involved, we had to develop a team-based working spirit with numerous meetings, formal note-taking, and structured decision making. This has helped when documenting everything in the data management plan. Not many other research teams are attempting to capture this type of music-related data. That is also why we believe sharing the dataset and the knowledge we have gained is essential.

Open Research is becoming the norm and is requested by many institutions and funders. However, while we welcome such a top-down policy change, there are many unsolved problems we have had to face when trying to implement this policy in real-life research. For example, even though some templates for DMPs exist, these are often developed as a desktop exercise and have not been tested on actual data captures. For relatively large projects like ours, including complex and contextual human data and involving several research teams from different institutions, writing a DMP is a complex procedure. Our approach is learning by doing. The MusicLab research concert series, of which MusicLab Copenhagen is the most ambitious to date, is an innovation project on its own. We try and fail and then try to improve based on our failures.

We did many things right at MusicLab Copenhagen. However, there are still numerous issues we have been unable to work through: data loss, poor recordings, unexpected privacy issues, expected copyright problems (see Sørbø et al., 2023), and many pre-processing challenges. Yet, the MusicLab Copenhagen Dataset is unique and rich and has already proven valuable for research, not least as evidenced by the articles that have already been published and others in the making. By sharing the dataset openly, we hope many others can also benefit from it and be inspired to share their data in the future.

Acknowledgments

Action Editor

Ian Cross, Faculty of Music, University of Cambridge.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

This project was approved by the Norwegian Centre for Research Data (NSD), reference number 915228.

Funding

ORCID iDs

Simon Høffding https://orcid.org/0000-0002-9739-9454
Laura Bishop https://orcid.org/0000-0002-0656-3969
Mathew Good https://orcid.org/0000-0003-3811-8920
Niels Chr. Hansen https://orcid.org/0000-0003-2142-6484
Live Håndlykken Kvale https://orcid.org/0000-0001-5264-0382
Eigil Lippert https://orcid.org/0000-0002-8234-5584
Nanette Nielsen https://orcid.org/0000-0002-2282-505X
Dana Swarbrick https://orcid.org/0000-0001-7242-9138
Finn Upham https://orcid.org/0000-0002-9948-2024
Atilla Vrasdonk https://orcid.org/0009-0007-7321-3276
Jonna Vuoskoski https://orcid.org/0000-0003-0049-4373
Sebastian Wallot https://orcid.org/0000-0002-3626-3940
Wenbo Yi https://orcid.org/0000-0003-1210-5312
Anne Danielsen https://orcid.org/0000-0003-1371-1533
Alexander Refsum Jensenius https://orcid.org/0000-0001-6171-8743

Notes

1. https://osf.io/v9wa4/
2. https://osf.io/fq8dr/files/osfstorage/ 620cd30211da1c03aff56979
3. https://www.youtube.com/watch?v=S4UVJybA6ZQ&t= 7829s
4. https://osf.io/yvbz8
5. https://www.uio.no/ritmo/english/projects/musiclab/2020/ musictestlab/
6. https://github.com/fourMs/MoMoCapture
7. https://github.com/finn42/MusicLabCopenhegan_MotionAlignment https://github.com/dana-and-monsters/Collectively-Classical. See also Høffding et al., 2021.

References

Bishop, L., González Sánchez, V., Laeng, B., Jensenius, A. R., & Høffding, S. (2021). Move like everyone is watching: Socialcontext affects head motion and gaze in string quartet performance. *Journal of New Music Research*, *50*(4), 392–412. https://doi.org/10.1080/09298215.2021.1977338

Bishop, L., Høffding, S., Lartillot, O., & Laeng, B. (2023, January 1). Mental effort and expressive interaction in expert and student string quartet performance. *Music & Science*, (6), 20592043231208000. https://doi.org/10.1177/20592043231208000

Danielsen, A., Paulsrud, T. S., & Hansen, N. C. (2023). "MusicLab Copenhagen": The gains and challenges of radically interdisciplinary concert research. *Music & Science*, (6). https://doi.org/10.1177/20592043231194747

Høffding, S., Bergstrøm, R. J. F., Bishop, L., Bravo, P. L., Burnim, K., Cancino-Chacón, C., Clim, A., Good, M.,

Hansen, N. C., Karlsen, E. S., Laeng, B., Lartillot, O., Lippert, E., Martin, R., Nielsen, N., Nørgaard, A., Omprakash, R., Rosas, F., Sjölin, F., Swarbrick, D., Sørbø, S., Sørensen, R. T., Upham, F., Vrasdonk, A., Vuoskoski, J., Yi, W., Øland, F., & Jensenius, A. R. (2021). MusicLab Copenhagen Dataset. Open Science Framework.

Høffding, S., Hansen, N. C., & Jensenius, A. R. (2024). Music research "in the wild" – Introducing the MusicLab Copenhagen Special Collection. *Music & Science*, (7). https://doi.org/10.1177/20592043241294161

Huvila, I., Andersson, L., & Sköld, O. (Eds.) (2024). *Perspectives on Paradata: Research and Practice of Documenting Process Knowledge*. Springer Nature.

Jensenius, A. R. (2021). Best versus good enough practices for open music research. *Empirical Musicology Review*, ISSN 1559-5749. *16*(1), 5–15. https://doi.org/10.18061/emr.v16i1.7646

Moss, F. C., & Neuwirth, M. (2021). FAIR, open, linked: Introducing the special issue on open science in musicology. *Empirical Musicology Review*, *16*(1), 1–4. https://doi.org/10.18061/emr.v16i1.8246

Sørbø, S., Good, M., Bergstrøm, R. J. F., & Kvale, L. H. (2023). The Library's Changing Role in Supporting Music Research: Part of the special collection on MusicLab Copenhagen. *Music & Science*, (6). https://doi.org/10.1177/20592043231186872

Upham, F. (2023). Using metrically-entrained tapping to align Mobile phone sensor measurements from in-person and live-stream concert attendees. In: Proceedings of The 17th International Conference on Music Perception and Cognition and The 7th Conference of the Asia-Pacific Society for the Cognitive Sciences of Music, edited by M. Sadakata: 357–361, https://www.uio.no/ritmo/english/people/postdoctoral-fellows/finnu/upham2023_icmpc_synchtap.pdf

Walters, W. H. (2020). Data journals: Incentivizing data access and documentation within the scholarly communication system. *Insights: The UKSG Journal*, *33*(1), 18. https://doi.org/10.1629/uksg.510

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Boiten, J.-W., Silva Santos, L. B. d., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, Article 160018. https://doi.org/10.1038/sdata.2016.18