German Medical Data Sciences 2023 — Science. Close to People.
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI230687

odML-Tables as a Metadata Standard in Microneurography

Alina TROGLIO[a,1], Aidan NICKERSON[b,d], Fabian SCHLEBUSCH[a], Rainer RÖHRIG[c], James DUNHAM[b], Barbara NAMER[a,2], and Ekaterina KUTAFINA[c,2]

[a] *Research Group Neuroscience, IZKF, Department of Physiology, RWTH Aachen University, Aachen, Germany*
[b] *School of Physiology, Pharmacology and Neuroscience Bristol Anaesthesia, Critical Care and Pain Research, University of Bristol, UK*
[c] *Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[d] *Eli Lilly and Company, Bracknell, UK.*
ORCiD ID: Alina Troglio, 0000-0001-7770-4389

Abstract. Metadata is essential for handling medical data according to FAIR principles. Standards are well-established for many types of electrophysiological methods but are still lacking for microneurographic recordings of peripheral sensory nerve fibers in humans. Developing a new concept to enhance laboratory workflows is a complex process. We propose a standard for structuring and storing microneurography metadata based on odML and odML-tables. Further, we present an extension to the odML-tables GUI that enables user-friendly search functionality of the database. With our open-source repository, we encourage other microneurography labs to incorporate odML-based metadata into their experimental routines.

Keywords. Metadata, FAIR, open metadata Markup Language (odML), microneurography, electrophysiology, pain

1. Introduction

1.1. Background

One of the greatest obstacles to scientific collaboration is the difficulty in sharing and understanding data between research groups. The efficient collection, storage, and analysis of experimental data require an organized data infrastructure. Furthermore, consistent metadata for complex electrophysiological experiments is essential for enabling FAIR principles [1].

One of the challenges is to strike a balance between making data FAIR and protecting confidentiality. In a recent study, Martorana et al. [2] examined how

[1] Corresponding Author: Alina Troglio, alina.troglio@rwth-aachen.de, Pauwelsstr. 30, 52074 Aachen
[2] Both authors contributed equally to the paper.

researchers have approached this challenge. The authors emphasized the critical role of metadata in restricted data flows.

Here, we present a solution for managing metadata based on odML and odML-tables for microneurography (see [3,4] for more details), a unique electrophysiological method for observing the activity of single peripheral nerve fibers in awake healthy humans and patients with neuropathy, chronic pain or itch [5,6]. In the field of microneurography, each direct neural recording is very valuable, particularly from patients with diseases caused by rare genetic variants. To maximize understanding of pathological changes, large data sets are required, which need to also include healthy comparators. Practically, such data can only be created with multi-center studies; therefore, solutions to facilitate data sharing and analysis are crucial. There exists a variety of approaches and structures for capturing experimental metadata, ranging from handwritten lab books to complex file and metadata schemas.

In microneurography, the lack of standardized metadata causes significant challenges not just for sharing of data, but also for effective searching and retrieving of relevant experimental data within research groups. Thus, there is a need for innovative approaches to handle both data and metadata in microneurography labs.

1.2. Objective and Requirements

The objective of this work is to determine the metadata requirements in microneurography and implement them in a way that allows efficient management of data and metadata.

We assembled the necessary requirements through an iterative discussion with experts who conduct microneurography experiments.

Individual experiments need to be structured and the separation of metadata and raw electrophysiology data files is essential, as metadata may contain personal information. The separation of raw data and metadata gives more control and restrictions, facilitating data sharing between research sites.

The integration of a metadata standard into the daily routine should be as comfortable as possible to ensure acceptance by the laboratory research team. The possibility of using standard office software (e.g., Microsoft Office) for documenting metadata is preferential.

In microneurography, the experimental protocols are often dynamically adjusted to the volunteer's individual characteristics, such as pain threshold, signal-to-noise or types of nerve fibers in a recording, etc. Thus, each individual experimental setup needs a separate template to structure the metadata. Due to the specifics of some acquisition systems (e.g., Dapsys, www.dapsys.net) and resulting time limits per recording, a single experimental session may be further broken down into several sub-sessions recorded as individual files. The metadata files should maintain flexibility so that each lab can add missing properties, however, it is important that the most relevant information is included in the proposed templates, that only minor changes are required and that sharing between labs is still possible.

Lastly, the files should be integrated into a complete database that is searchable, allowing for specific files to be retrieved based on parameters such as stimulation protocol, fiber types, patients' diagnosis or other variables.

2. State of the art

2.1. Related Work

The task of organizing and describing scientific experiments is a significant challenge. The Investigation-Study-Assay (ISA) Abstract Model [7] is a metadata framework addressing this task. The goal is to simplify the exchange and management of research data across domains using metadata of the three (investigation, study, and assay) components.

The development of a data-handling infrastructure in the field of electrophysiology has also been approached through various methods in recent years. Mouček et al. [8] worked on a complete infrastructure for research in electrophysiology with several software and hardware components. They collect data and metadata in a web-based portal that makes retrieving and analyzing electrophysiological data more suitable and efficient.

Rübel et al. [9] presented the Neurodata Without Borders (NWB) data language which provides a standardized structure for combining data and metadata to facilitate sharing and storage of experiments. Software is provided that facilitates the research on neurophysiological data.

Generic metadata standards, such as DataCite [10] and Dublin Core [11] offer domain-independent solutions for describing data, thereby ensuring that data is accessible and reusable. Dublin Core is a metadata standard for digital resources in general. DataCite is more specific for research data by including, for example, funding information and access rights and it provides a persistent identifier for data sets, which makes citation possible.

Zehl et al. [12] addressed strategies for the management of metadata in the specific setting of neurophysiological laboratories. It is emphasized that electrophysiological experiments involve multiple hardware and software components, and varying lab cultures and practices, which makes this field and reproducibility so complex. As a solution, the authors propose the work of Grewe et al. [13], who designed odML (open metadata Markup Language) for metadata in neurophysiology. odML is both human- and machine-readable and builds a hierarchical structure to define entries as key-value pairs, but the use of the language requires programming skills. This requirement often cannot be met in small research groups in medical research communities, which is why Sprenger et al. [14] introduced odML-tables. This solution simplifies the usage of odML by providing a graphical user interface (GUI) and a script that converts the odML tree structure into a flat table format, e.g., csv or xls.

2.2. Shortcomings

Currently, there exist methodologies for handling metadata in electrophysiology but to date there is no generally standardized approach in microneurography. We require machine-readable structures that are specifically adjusted to the experimental flow of microneurography. To proceed, we need to select a solution and adapt it to our data, followed by a test to assess its compatibility with the laboratory routines.

3. Concept

Based on the requirements and available solutions for electrophysiology, we decided to proceed with odML and odML-tables. This option enables the collection of metadata for an individual experiment in an Excel table and this can be easily fitted into the lab flow. Further, individual files can be aggregated into a database and stored separately from the electrophysiological data. We have created two table-based templates that can be converted to the odML format using the odML-tables GUI. Thus, we have a simple, flexible, and lightweight solution. The use of metadata templates for data collection is a well-established concept that allows a flexible structure for different laboratories and manual changes to the metadata. Using odML-tables facilitates odML for researchers without programming skills but is still machine-readable. Additionally, odML-tables provides several functionalities, for example, merging multiple files or filtering the metadata collection. Finally, odML is an open-source project with an active community, which supports the ongoing development.

4. Implementation

4.1. Solution / Results

Metadata templates The first draft of two templates, one for general experimental metadata and the other one for individual data recordings were defined at the microneurography lab in Aachen and iteratively discussed with the lab in Bristol. Eventually, the final version was agreed upon. To evaluate the solution, the stakeholders in Aachen and Bristol installed the software, created the database, and conducted a search. Their feedback is collected to refine and improve the solution.

The experimental template covers general information about the experiment, recording equipment, and the subject. General information may include the laboratory, the experimenter, or the recording date. For the recording equipment, the hardware information, such as electrode types and acquisition system. The section about the experimental subject describes medically important data such as age, sex, and any medications.

To describe the actual content of the recording file, the recording template was created. The recording template is divided into 7 subcategories: *General, Fibers, ElectricalStimulus, MarkingMethod, MechanicalStimulus, HeatStimulus,* and *Chemicals*. The *Fibers* category describes, for example, how many neural fibers have been observed during the experiment in this recording and what type of fiber they are (e.g., mechano-sensitive C-fiber). There is also a marker that indicates the occurrence of spontaneous activity which is linked to neuropathic pain [15]. Additionally, the stimulation protocol (electrical, mechanical, thermal, or chemical) is included. The marking method (see [16] for more details) is used to track the action potentials of active fibers to electrical background pulses and identify their responses to different types of stimulation semi-quantitatively. It is possible to state if the marking method was used (yes/no) and the frequency of the background pulses, which is important information since the frequency of this stimulation changes the excitability of the nerve fibers. Figure 1 shows the hierarchical structure of the metadata.

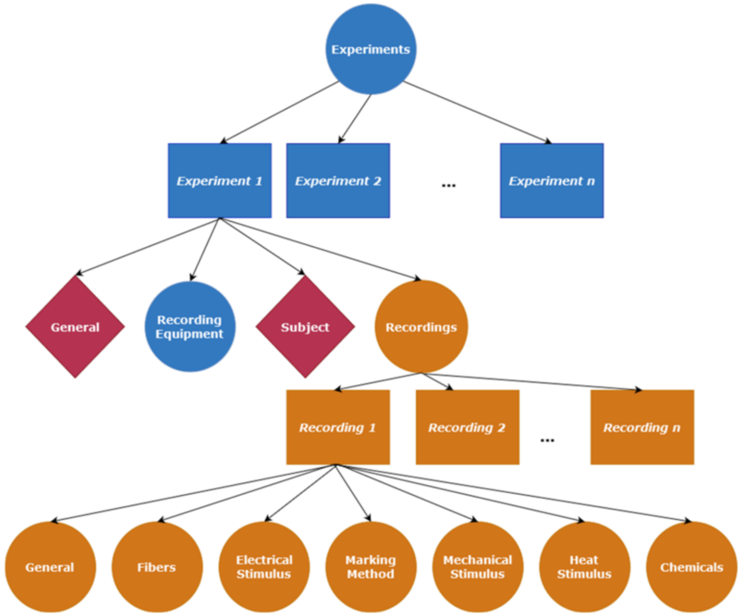

Figure 1. Hierarchical structure of experiment template (blue) and recording template (orange) in the complete metadata collection. The sections in the red diamonds may contain quasi-identifiers, such as the date of the experiment.

4.2. System in Use

Data exchange pipeline Our odML-based metadata standard and our software [17] for handling raw data, which is developed in parallel, are effectively integrated into the data workflow (see Figure 2). This approach enables us to separate the storage of data and metadata into distinct databases and restricts access to confidential patient information. Using the odML-tables GUI simplifies converting experimental metadata from a standard table-based representation to the more complex odML format for collection purposes, resulting in a more convenient experience. Maintaining distinct metadata databases with and without quasi-identifiers ensures a FAIR data infrastructure to the extent that data as well as metadata can be shared with other microneurography laboratories.

Integration into the daily lab routine In order to incorporate odML into the lab routine, guidelines for use and installation were created to support the installation of Python and all required packages. The code, the templates and the guidelines can be found in our GitHub repository[3]. Experiment templates and recording templates were completed for all recent microneurography sessions in the Aachen lab, followed by the conversion of the filled-out templates from the table to the odML format using the provided odML-tables functionality. All odML files were then merged and aggregated into a single odML file that serves as a complete database for metadata. The odML-tables GUI provides five functionalities, namely converting between table and odML format,

[3] https://github.com/Digital-C-Fiber/odMLtablesForMNG

creating a new table, comparing entries within a single odML file, merging contents of odML files, and filtering contents of odML files to clean up a database.

Search functionality Historically, there has been a lack of a search function adapted to meet the specific needs of microneurography.

We implemented search functionalities to the GUI to perform simple and straightforward searches for property-value pairs in the graphical user interface (see Figure 3 for illustration). We developed a search dialog (wizard) that supports two search options: free text search and pairs of property-value combinations. First, the complete metadata collection is loaded into the tool. Then, the user can choose one of the two options.

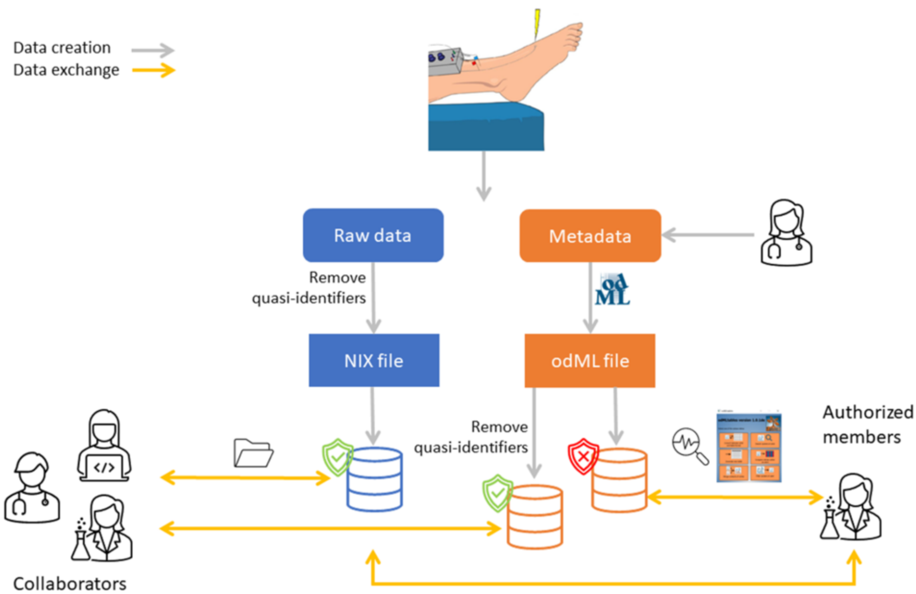

Figure 2. Overview of the data sharing process. Initially, microneurography experiments generate raw data and metadata, followed by the removal of quasi-identifiers (e.g., date of recording). The pseudoanonymized raw data is then stored as a NIX file. Researchers complete the designed metadata templates and convert them to odML format. To maintain data confidentiality, we have two odML files, one that can be shared with other microneurography labs after all quasi-identifiers were removed and one for internal use only. To support data retrieval, the odML-tables GUI incorporates a novel search function.

The properties are read directly from the template. Our implementation allows up to three combinations, e.g., the user can find all recordings containing only a single mechano-insensitive C-fiber by using the combination "FibreCount": "1" and "FibreTypes": "CMi". The search integration into the GUI makes the search very user-friendly for an interdisciplinary team with different levels of programming skills.

Current state A collection of 36 odML files has been generated within the Aachen laboratory for all recent experiments and merged into a single odML database file. Currently, the testing phase has started in the Bristol lab.

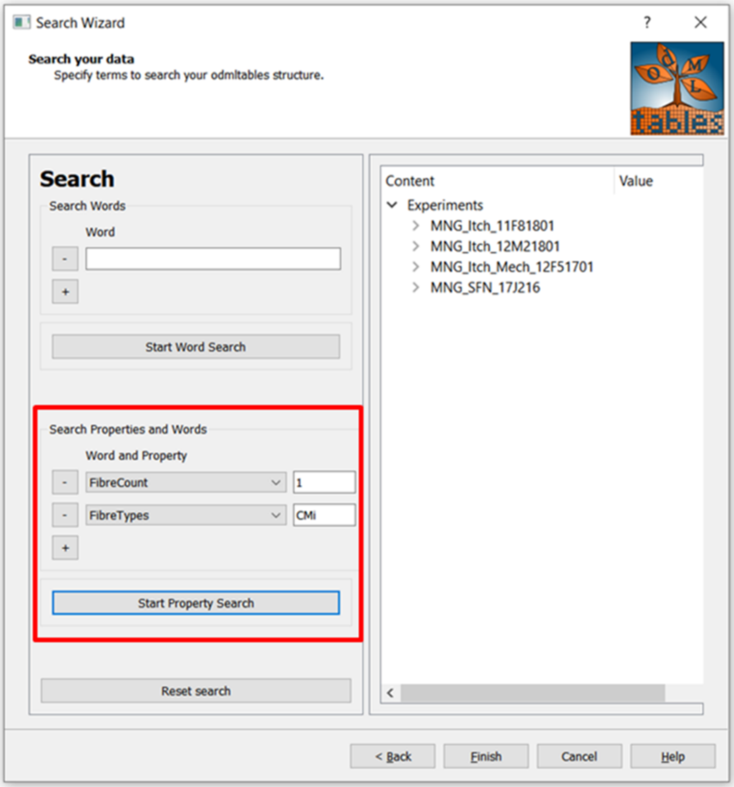

Figure 3. Exemplary search. This is an example of a search conducted using the search wizard, our extension of the existing odML-tables GUI. We are searching for recordings that contain a single "CMi" C-fiber.

5. Lessons learned

OdML-tables provides an optimal bottom-up solution for microneurography metadata handling. One of the main reasons is that it allows for easy separation between the anonymous (after controlled conversion from a proprietary format to e.g., NIX) raw data and potentially sensitive experimental information.

Another point is the possibility of using simple, human-readable, and microneurography-specific Excel tables which can be converted to odML and integrated into the hierarchical database with a user-friendly GUI. Other solutions, such as the web-based portal developed by Mouček et. al. [8] and NWB [9] provide much wider options and functionalities, but they come with the cost of more complex adaptation processes and lower expected user acceptance in microneurography labs. The same applies to international metadata standards, e.g., DataCite.

Through the international collaboration of several stakeholders, we ensure that the templates contain the most relevant information. However, any missing properties can be effortlessly incorporated into the templates at a later stage, which provides flexibility. The current version of the odML-tables GUI and the searchable database already support our work on computational modeling of C-fibers (e.g., [18]), by efficiently finding experiments based on domain-specific criteria. It is also being integrated into the next version of the software platform for microneurography analysis “openMNGlab” [17].

In summary, the implementation fulfilled the specification and has been successfully incorporated in the lab. Although the individual components of the solution were available for electrophysiology, adjustments to a specific type of experiment and specific lab required novel solutions and allocated resources and should be treated as an important part of the standard research process.

6. Conclusion

We have developed an odML-based candidate for a microneurography metadata standard through the iterative collaborative work of two laboratories. The metadata templates have been incorporated into the routine lab work in Aachen and integrated into an odML database. In response to user feedback, the existing odML-tables GUI was enhanced with new search functionality. Our ultimate goal is to have FAIR-compliant microneurography data.

In the next step, we will confirm the generalization of the odML-tables solution with microneurography data collected at the University of Bristol. We will iteratively improve the format and GUI through discussions with other microneurography labs. The adoption of a common standard, which additionally allows full control of the potentially sensitive information, will enable the creation of valuable databases of retrospective and prospective records in which patients with different neurological conditions are well represented.

Declarations

Ethical vote: N/A

Conflict of Interest: AN is a current or former employee of Eli Lilly and Company Inc., and may own stock in this company.

Contributions of the authors: BN, EK and RR conceived the concept; BN, JD and AN consulted the work from the user perspective; FS, EK and AT developed the solution and the templates; FS and AT implemented the solution; EK, BN, JD and RR supervised the development; AT drafted the paper and all authors substantially revised. All authors approved the final version and agreed to be accountable for all aspects of the work in ensuring that questions related to accuracy.

Acknowledgments

This work was supported by the Excellence Initiative of the German Federal and State Governments G:(DE-82)EXS-SF-SFDdM013, a grant by the Interdisciplinary Center for Clinical Research IZKF TN1-6/IA 532006 and by the DFG grant NA 970 7-2 and 6-2. BN is supported by a grant from the Interdisciplinary Center for Clinical Research within the faculty of Medicine at the RWTH Aachen University.

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 15. March 2016;3:160018.
[2] Martorana M, Kuhn T, Siebes R, van Ossenbruggen J. Aligning restricted access data with FAIR: a systematic review. PeerJ Comput Sci. 2022;8:e1038.
[3] Vallbo AB, Hagbarth KE. Activity from skin mechanoreceptors recorded percutaneously in awake human subjects. Exp Neurol. July 1968;21(3):270–89.
[4] Torebjörk HE, Hallin RG. Responses in human A and C fibres to repeated electrical intradermal stimulation. J Neurol Neurosurg Psychiatry. June 1974;37(6):653–64.
[5] Ackerley R, Watkins RH. Microneurography as a tool to study the function of individual C-fiber afferents in humans: responses from nociceptors, thermoreceptors, and mechanoreceptors. J Neurophysiol. 1. December 2018;120(6):2834–46.
[6] Kutafina E, Becker S, Namer B. Measuring pain and nociception: Through the glasses of a computational scientist. Transdisciplinary overview of methods. Frontiers in Network Physiology. 2023;3.
[7] Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, et al. Toward interoperable bioscience data. Nat Genet. February 2012;44(2):121–6.
[8] Mouček R, Ježek P, Vařeka L, Rondík T, Brůha P, Papež V, et al. Software and hardware infrastructure for research in electrophysiology. Front Neuroinform. 2014;8:20.
[9] Rübel O, Tritt A, Ly R, Dichter BK, Ghosh S, Niu L, et al. The Neurodata Without Borders ecosystem for neurophysiological data science. Colgin LL, Jadhav SP, Herausgeber. eLife. 4. October 2022;11:e78362.
[10] Brase J. DataCite - A Global Registration Agency for Research Data. In: 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology. 2009. S. 257–61.
[11] Dublin Core Metadata Element Set, Version 1.1: Reference Description. https://www.dublincore.org/specifications/dublin-core/dces/.
[12] Zehl L, Jaillet F, Stoewer A, Grewe J, Sobolev A, Wachtler T, et al. Handling Metadata in a Neurophysiology Laboratory. Frontiers in Neuroinformatics. 2016;10.
[13] Grewe J, Wachtler T, Benda J. A Bottom-up Approach to Data Annotation in Neurophysiology. Frontiers in Neuroinformatics. 2011;5.
[14] Sprenger J, Zehl L, Pick J, Sonntag M, Grewe J, Wachtler T, et al. odMLtables: A User-Friendly Approach for Managing Metadata of Neurophysiological Experiments. Frontiers in Neuroinformatics. 2019;13.
[15] Kleggetveit IP, Namer B, Schmidt R, Helås T, Rückel M, Ørstavik K, et al. High spontaneous activity of C-nociceptors in painful polyneuropathy. Pain. October 2012;153(10):2040–7.
[16] Schmelz M, Forster C, Schmidt R, Ringkamp M, Handwerker HO, Torebjörk HE. Delayed responses to electrical stimuli reflect C-fiber responsiveness in human microneurography. Exp Brain Res. 1995;104(2):331–6.
[17] Schlebusch F, Kehrein F, Röhrig R, Namer B, Kutafina E. openMNGlab: Data Analysis Framework for Microneurography - A Technical Report. Stud Health Technol Inform. 21. September 2021;283:165–71.
[18] Kutafina E, Troglio A, de Col R, Röhrig R, Rossmanith P, Namer B. Decoding Neuropathic Pain: Can We Predict Fluctuations of Propagation Speed in Stimulated Peripheral Nerve? Frontiers in Computational Neuroscience. 2022;16.