OPEN ACCESS

Short Report

Research data management – chances, challenges and solutions for audiology

Forschungsdatenmanagement – Chancen, Herausforderungen und Lösungen für die Audiologie

Abstract

This short report addresses relevant aspects of sharing, pooling and publishing audiological research data. We discuss legal frameworks and technical and organizational solutions. We introduce the Oldenburg Hearing Health Record (OHHR) as an example on data re-organization, anonymization and tooling for the open access publication is described. Federated learning is discussed as an alternative if regulatory or ethical obstacles prevent direct pooling of data from different sources.

Keywords: research data management, data pooling, data exchange, legal frameworks, open access publication, federated learning

Zusammenfassung

Dieser Kurzbeitrag befasst sich mit relevanten Aspekten der gemeinsamen Nutzung, Zusammenführung und Veröffentlichung audiologischer Forschungsdaten. Es werden rechtliche Rahmenbedingungen sowie technische und organisatorische Lösungen erörtert. Wir stellen das Oldenburg Hearing Health Record (OHHR) als Beispiel für die Reorganisation und Anonymisierung von Daten vor und beschreiben Werkzeuge für die Open-Access-Veröffentlichung. Föderiertes Lernen wird als eine Alternative diskutiert, wenn regulatorische oder ethische Hindernisse die direkte Zusammenführung von Daten aus verschiedenen Quellen verhindern.

Daniel Berg[1]
Mareike Buhl[2]
Dagmar Krefting[3]

1 Carl von Ossietzky University Oldenburg, Department of Medical Physics and Acoustics, Cluster of Excellence Hearing4all, Oldenburg, Germany

2 Université Paris Cité, Institut Pasteur, AP-HP, INSERM, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, Paris, France

3 University Medical Center Göttingen, Department of Medical Informatics and Campus Institute Data Science, Göttingen, Germany

Introduction

In recent years, research data management (RDM) according to the FAIR principles [1] throughout the entire data lifecycle — from collection to publication and reuse — has gained increasing importance across all fields of research. The reasons for this are manifold. Firstly, the desire to enhance the transparency and reproducibility of research is a response to the reproducibility crisis [2]. This is reflected in an increase in external requirements. Many research funding institutions now require a detailed research data management plan at the application stage, and most publishers demand that the data underlying a scientific publication be made available. Secondly, publishing research data can increase its reach and thus enhance a researcher's reputation. On the other hand, publishing research data can increase the reach of research and thus enhance the reputation of a researcher. Finally, many current research methods such as machine learning and artificial intelligence require large amounts of data (big data), so there are hardly any alternatives to reusing third-party data.

To boost reusability of data, it is crucial to enrich data with extensive machine readable metadata. Metadata, typically defined as information describing the actual data, may encompass a broad variety of structured information that is used to organize and discover data. The metadata should follow established community standards, and should be made findable for others, for example in a community-specific repository. This allows researchers to discover and identify data that are relevant and suitable or applicable for their own research.

For pooling data from different sources, it is furthermore essential to establish interoperability of the raw data. This can be achieved e.g. by using the same data format during collection or by converting the data to the same target format. Data standards enable interoperability by complete documentation of data and metadata [1], but to date only starting points of audiological data standards are available. One future candidate could be the openEHR

format [3]. In openEHR audiological standards (archetypes) are still missing, but an international research group at the European Federation of Audiology Societies (EFAS) is currently working on the definition of those standards [4].
This report highlights some legal issues concerning data sharing and pooling, presents the Oldenburg Hearing Health Record (OHHR) [5], [6] as an example of open access publication of audiological data, and describes the application of federated learning in audiology.

Exchange and pooling of data – regulatory frameworks and technical-organizational solutions

A common challenge in data usage is the integration of different data sources. Local information systems and data repositories typically only cover a small part of the big picture, that would be needed to gain insights about general mechanisms. Integrating health data from different data sources may follow two different paradigms — (a) record linkage to get a holistic view on the health status and development across different stations in the patient path and (b) data pooling, also called data federation, to get a broad view on the inter-individual variability. Record linkage may combine clinical documentation from out- and inpatient care with imaging, sensory, genetic or reimbursement data of a person. The integration of in- and outpatient data along the patient path is currently piloted by the Digihubs as part of the Medical informatics initiative [7]. Integration of cohort studies and health insurance data has been recently proposed by the Network University Medicine [8]. Data pooling may combine the same data type from different health institutions, registries and studies. However, both methods require exchange of data across institutional borders, facing several challenges. In particular regulatory aspects regarding data protection plays a critical role in biomedical research. In Europe, the General Data Protection Regulation (GDPR), prohibits usage of health data with article 9 — among other data types classified as sensitive — unless there is a legal basis at all. Some legal bases are for example explicit informed consent, medical diagnostics and care [9]. Further legal specifications encompass data collection due to the infection protection law, cancer and transplant registries and statutory health insurance regulations. Use of health data for research has therefore been typically made possible through the informed consent. Research with health data was in principle possible through article 89 of the GDPR, but has been rarely used at least in Germany. Data Protection officers typically argued that the barriers should be very high. The required proof that the intended research could not be realized with informed consent on one hand or anonymization, which would bring the data outside the scope of the GDPR on the other hand, had been considered in the research community as very difficult to convince data protection officers. Furthermore, in Germany data protection is regulated on the level of the states. The heterogeneity of the state's data protection regulations made it very difficult to share health data across the federal states [10]. While anonymization is a measure to enable reuse of health data without consent, this might be difficult to realize in practice. It requires that the data does not contain any "person-related" data and cannot be "reidentified". For high dimensional data with many data items, the proof of full anonymity is difficult to provide, and data protection officers regularly have questioned it [11]. Furthermore, record linkage is no longer possible with anonymized data, limiting the usefulness of such data sets.
However, the open science movement has demanded better accessibility of data used in research. So, in recent years, ways have been explored to fulfill this demand — by publication of anonymized data and to enable sharing of health data based on a consent that explicitly allows for usage in future research, the so-called broad consent. Anonymized open health data can be found from German initiatives, such as the National Pandemic Cohort and the research data center of the Federal Institute for Drugs and Medical Products [12]. Another way to use health data without consent is federated analysis, where only aggregated data is shared (see chapter 3, "Federated learning in audiology").
Broad consent has been pioneered by cooperative health research projects such as the German Center for Cardiovascular Research [13]. A major step forward was the approval of the broad consent for healthcare data, developed within the Medical informatics initiative, by the conference of the independent state's data protection officers in 2020. The consent allows transfer of pseudonymized — i.e. there exist a mapping of the pseudonym to identifying data, typically held by a trusted party — data for ethically reviewed research questions [14]. However, the implementation into the care process is complex and is currently mainly realized at university hospitals. Therefore, the German Health Data Usage Act (Gesundheitsdatennutzungsgesetz — GDNG) [15] is an important milestone, as it allows for research for healthcare institutions on pseudonymized data of their patients in terms of self-research, but also in funded cooperative research projects. The GDNG has become effective in March 2024. It has been developed with anticipation of the European Health Data Space (EHDS) that demands comprehensive health data sharing for research, policy and innovation within the European Union. The EHDS has become effective in March 2025 [16]. The members of the EU now have two years to define the implementation acts, and most data categories — from electronic health records to clinical studies — are planned to be available from March 2029 on.
With the recent developments exchange and pooling of health data in particular for research purposes are becoming easier to realize.

The OHHR (Oldenburg Hearing Health Record) as an example on data re-organization, anonymization and tooling

The OHHR provides a publicly accessible dataset that can be used to advance hearing health research. It includes hearing related data collected from 581 participants (aged 18–86 years; 255 female; mean age=67.31 years; standard deviation=11.93 years) between 2013 and 2015 at the Hörzentrum Oldenburg in collaboration with the Cluster of Excellence "Hearing4all".

The OHHR contains audiological measures (pure tone audiogram, categorical loudness scaling, speech-in-noise tests), multiple subjective questionnaires and a detailed expert interview (anamnesis), different cognitive measures and demographic information for each subject. Due to the large diversity of data from different domains, the dataset is particularly suitable for multimodal analysis.

The original data were collected in multiple Excel tables, each containing direct identifiers of the subjects and the raw data in many multimodal columns. To publish the dataset according to the FAIR principles the data have been prepared in several steps:

Anonymization: The General Data Protection Regulation (GDPR) requires informed consent for the publication of personal data. Since a signed consent form was not available for all subjects, the data was converted into non-personal data through anonymization in collaboration with a data protection officer. All direct identifiers were removed, demographic data were checked for a maximum number of contained indirect identifiers (e.g. locations, gender, religion) of three. Afterwards a K-anonymity of 4 was applied to the data [17]. This was achieved by omitting some data fields (e.g. year of birth was kept, month was removed) and grouping values into more vague categories (e.g. some net income groups were combined). Two subjects with unique characteristics had to be removed from the dataset. Finally, a unique anonymous identifier was assigned to each subject to allow the identification within the different tables of the dataset.

Data conversion/re-organization: To enable the reusability and interoperability of data, open data types and file formats should be preferred over proprietary and/or platform-dependent formats (e.g. EXCEL). Furthermore, the data structures should allow data aggregation and evaluation as well as import or conversion into different systems. For the OHHR the JSON [18] was used because it is at least partial type safe and offers the possibility of creating table definitions in schema files [19]. The data were re-organized into relational structured JSON data tables. These can be imported seamlessly into the openEHR format using the inbuilt ETL processes (Extract, Transform, Load) after the corresponding archetypes are available. Metadata were added in two formats: table definitions with descriptions of all data fields and allowed value ranges are stored in JSON schema files to allow the validation of data records against the corresponding schemas. Additionally, detailed descriptions of all tables and fields are available in RTF text format.

Finally, the publication of the dataset was supplemented by a collection of software tools to facilitate the reusability of the data: EXCEL tables demonstrate how to use JSON tables as data sources and define queries to join the related tables. MATLAB and Python scripts show examples of reading and joining the different JSON tables. A SQL tool script creates a complete database and imports all tables into the corresponding SQL tables.

The OHHR was published on Zenodo [5] under the Creative Commons License CC-BY [20] to grant accessibility and maximum flexibility for the reusage of the data. Zenodo was selected as repository in favour to other existing medical data repositories for two reasons: currently there is no particular repository available for audiological or other hearing related medical data. The cluster of excellence "Hearing4all" is planning to set up an "Open Dataspace for Audiology" in the upcoming funding period, which might close this gap in the future. Additionally, we think the reach of the Hearing4all community on Zenodo can increase the visibility of the data record considerably. Nevertheless, the decision to use a particular repository should always be reconsidered before publishing data sets.

This type of procedure can be particularly useful for inventory data for which no declaration of consent has been obtained, provided that this data can be anonymized.

Federated learning in audiology: supervised and unsupervised applications on audiological databases

Audiological data are typically collected in local, decentralized databases, exhibiting different data structures, choice of tests, or varying data quality. The same or similar audiological tests are conducted in clinical and research contexts. If these data were analyzed together, knowledge about many patients worldwide could be exploited to obtain a representative overview of existing patient patterns, and classifications into hearing loss or hearing device categories could support clinical decision-making. This can ultimately lead to a clinical decision-support system (CDSS) for audiology, as proposed by Buhl et al. [21]. To incorporate population-representative data in a CDSS or in general AI methods for knowledge extraction, these tools have to operate on large amounts of interoperable data. Federated learning is a suitable method if databases cannot be pooled.

Federated learning (FL) is an approach for training a machine learning model on data from distributed databases without the need for individual data exchange [22]. The general principle is to train a local model on every database ("the model comes to the data"), send learned parameters to a central server to combine the local models to a global model, and apply the global model on every database for (e.g.) classification [23]. Thereby, only aggregated data is shared.

FL methods are able to work on heterogeneous data sources due to properties such as dealing with different distributions or unbalanced data sets [23]. In audiology, this refers for example to including audiological data from hearing aid acousticians and cochlear implant databases, which typically contain different degrees of hearing loss and may have a different number of patient data available. Global models can be updated with additional local data sources and be retrained in several communication rounds to improve the global model [22].

Different medical disciplines successfully applied FL, e.g., for lung cancer tumor staging and survival based on data from eight institutions complying with the FAIR principles [24], and for mortality prediction for hospitalized Covid-19 patients based on electronic health record data from five institutions [25].

Towards integrating knowledge from databases for audiological decision-support, two studies will be presented that introduce FL to audiology, for unsupervised clustering and supervised classification.

Study 1: Auditory profiling for unsupervised characterization of audiological databases

Auditory profiles are a data-driven, unsupervised method to characterize patients in audiological databases by the combination of different audiological test results, such as audiogram, speech test, and loudness scaling [26]. The generation of profiles is based on model-based clustering, resulting in most distinct groups of patients according to their data. Saak et al. [27] developed an approach for merging the generated profiles obtained from several databases, while only considering information about distributions of the features per DB within the profiles, which corresponds to parameters of local models. The merged profile set corresponds to the global model. It was shown that the resulting profiles from two databases (from Hörzentrum Oldenburg, one of them being the OHHR [5]) plausibly extend the previous set of profiles, e.g., by capturing more severe hearing loss patterns in additional profiles. This principle can be expanded to more databases to eventually achieve a population-representative profile set. If databases contain different features (audiological tests), the merging step is performed based on the common features, and knowledge from additional, DB-specific features is maintained as distributions per profile. To show that the AP approach quantifies to a full FL approach, future work has to disentangle the influence of common and DB-specific features in a classification into profiles.

Study 2: Kernel-based federated learning for supervised classification in hearing loss categories

For a supervised classification based on several audiological databases, i.e., into hearing loss or hearing device categories, a federated learning approach needs to capture the properties of the databases, and include a suitable and interpretable design of features. Kernels provide the potential to optimize the features from audiological databases by projecting the features in a higher-dimensional vector space [28]. Kernel matrices can be derived for categories contained in databases, and flexibly combined by linear combination [29]. Such a framework can then be used to establish classification on single, as well as combined databases in a federated learning context.

Auditory profiling and kernel-based federated classification contribute to characterizing the hearing-impaired population based on decentralized audiological databases, which contain important statistical knowledge that can be exploited to support audiological diagnostic decisions. Both approaches will provide important explainable learnings about differences between databases, even if data can be pooled.

Summary and conclusion

Research data management has become increasingly important over the last decade, and notable progress in health data sharing and reuse on regulatory and technical-organizational level has been made in Germany and Europe.

The OHHR provides a good example of how auditory data can be made available for research and reuse according to the FAIR principles and can serve as a best practice example, on open data, providing a good structure without yet relating to a standard (later possible due to extensive metadata already provided in the OHHR dataset).

Federated learning provides an alternative to pooling if this is not possible, but also content-related advantages to learn about individual data and relationships between databases.

Taken together, improving RDM and sharing of research data provide important research opportunities for audiology, for example, related to data-driven audiological decision-support.

The growing RDM community as well as extensive national and international funding programs on data management support the efforts to improve RDM.

However, as for now the field of audiology seems to be underrepresented in the relevant initiatives and consortia such as NFDI4Health. Thus it will be important in the future to better link audiological research with these initiatives.

Notes

Acknowledgments

This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 – Project ID 390895286.
Mareike Buhl was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 496819293, and supported by a grant from Fondation Pour l'Audition (FPA) to the CERIAH (FPA IDA10). This work has benefited from a French government grant managed by the Agence Nationale de la Recherche under the France 2030 program, reference ANR-23-IAHU-0003.
Part of the work is based on the Medical Informatics Initiative and the Network University Medicine, both funded by the BMFTR – Federal Ministry of Research, Technology and Space (01ZZ2324A, 01ZZ2103A, 01KX2121).

Conference presentation

This contribution was presented at the 27th Annual Conference of the German Society of Audiology and published as an abstract [30].

Competing interests

The authors declare that they have no competing interests.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. DOI: 10.1038/sdata.2016.18
2. Stupple A, Singerman D, Celi LA. The reproducibility crisis in the age of digital medicine. NPJ Digit Med. 2019 Jan 29;2:2. DOI: 10.1038/s41746-019-0079-z
3. openehr.org. openEHR International (CIC). [cited 2025 Apr 17]. Available from: https://openehr.org/
4. EFAS Working Group on Data formats for Big Data in Audiology. [cited 2025 Apr 28]. Available from: https://efas.ws/data-formats-for-big-data-in-audiology/
5. Jafri S, Berg D, Buhl M, Vormann M, Saak S, Wagener KC, Thiel C, Hildebrandt A, Kollmeier Bl. OHHR – The Oldenburg Hearing Health Record [Dataset]. Zenodo; 2025. DOI: 10.5281/zenodo.14177902
6. Jafri S, Berg D, Buhl M, Vormann M, Saak S, Wagener KC, Thiel C, Hildebrandt A, Kollmeier B. The Oldenburg Hearing Health Record (OHHR). Sci Data. 2025;12:1546. DOI: 10.1038/s41597-025-05884-y
7. Krefting D, Bavendiek U, Fischer J, Marx G, Molinnus D, Panholzer T, Prokosch HU, Leb I, Weidner J, Sedlmayr M. Die digitalen Fortschrittshubs Gesundheit – Gemeinsame Datennutzung über die Universitätsmedizin hinaus [Digital health progress hubs-data integration beyond university hospitals]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2024 Jun;67(6):701-9. DOI: 10.1007/s00103-024-03883-9
8. Schmitt J, Ihle P, Schoffer O, Reese JP, Ortmann S, Swart E, Hanß S, Hoffmann F, Stallmann C, Kraus M, Semler SC, Heyder R, Vehreschild JJ, Heuschmann P, Krefting D, Sedlmayr M, Hoffmann W; und die gemeinsame Arbeitsgruppe „Externe Daten" des Netzwerks Universitätsmedizin (NUM) und der Medizininformatik-Initiative*. Datennutzung für eine bessere Gesundheitsversorgung–Plädoyer für eine kooperative Forschungsdatenplattform der gesetzlichen und privaten Krankenversicherung und dem Netzwerk Universitätsmedizin (NUM) [Access to and use of Data for better Healthcare: A Plea for a cooperative data and Research Infrastructure of Statutory and Private Health Insurers and the Network University Medicine (NUM)]. Gesundheitswesen. 2025; 87(S 02):S279-S288. DOI: 10.1055/a-2438-0670
9. European Commission. Regulation - 2016/679 - EN - gdpr - EUR-Lex. [cited 2025 Apr 19]. Available from: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
10. Molnár-Gábor F, Sellner J, Pagil S, Slokenberga S, Tzortzatou-Nanopoulou O, Nyström K. Harmonization after the GDPR? Divergences in the rules for genetic and health data sharing in four member states and ways to overcome them by EU measures: Insights from Germany, Greece, Latvia and Sweden. Semin Cancer Biol. 2022 Sep;84:271-83. DOI: 10.1016/j.semcancer.2021.12.001
11. Olatunji IE, Rauch J, Katzensteiner M, Khosla M. A Review of Anonymization for Healthcare Data. Big Data. 2024 Dec;12(6):538-55. DOI: 10.1089/big.2021.0169
12. Koll CEM, Hopff SM, Meurers T, Lee CH, Kohls M, Stellbrink C, Thibeault C, Reinke L, Steinbrecher S, Schreiber S, Mitrov L, Frank S, Miljukov O, Erber J, Hellmuth JC, Reese JP, Steinbeis F, Bahmer T, Hagen M, Meybohm P, Hansch S, Vadász I, Krist L, Jiru-Hillmann S, Prasser F, Vehreschild JJ; NAPKON Study Group. Statistical biases due to anonymization evaluated in an open clinical dataset from COVID-19 patients. Sci Data. 2022 Dec 21;9(1):776. DOI: 10.1038/s41597-022-01669-9
13. Hoffmann J, Hanß S, Kraus M, Schaller J, Schäfer C, Stahl D, Anker SD, Anton G, Bahls T, Blankenberg S, Blumentritt A, Boldt LH, Cordes S, Desch S, Doehner W, Dörr M, Edelmann F, Eitel I, Endres M, Engelhardt S, Erdmann J, Eulenburg K, Falk V, Felix SB, Frank D, Franke T, Frey N, Friede T, Geidel L, Germans L, Grabmaier U, Halle M, Hausleiter J, Jakobi V, Jebran AF, Jobs A, Kääb S, Karakas M, Katus HA, Klatt A, Knosalla C, Krebser J, Landmesser U, Lee M, Lehnert K, Lesser S, Leyh K, Lorbeer R, Mach-Kolb S, Meder B, Nagel E, Nolte CH, Parwani AS, Petersmann A, Puls M, Rau H, Reiser M, Rienhoff O, Scharfe T, Schattschneider M, Scheel H, Schnabel RB, Schuster A, Schmitt B, Seidler T, Seiffert M, Stähli BE, Stas A, J Stocker T, von Stülpnagel L, Thiele H, Wachter R, Wakili R, Weis T, Weitmann K, Wichmann HE, Wild P, Zeller T, Hoffmann W, Zeisberg EM, Zimmermann WH, Krefting D, Kühne T, Peters A, Hasenfuß G, Massberg S, Sommer T, Dimmeler S, Eschenhagen T, Nauck M. The DZHK research platform: maximisation of scientific value by enabling access to health data and biological samples collected in cardiovascular clinical studies. Clin Res Cardiol. 2023 Jul;112(7):923-41. DOI: 10.1007/s00392-023-02177-5

14. Zenker S, Strech D, Jahns R, Müller G, Prasser F, Schickhardt C, Schmidt G, Semler SC, Winkler E, Drepper J. National standardisierter Broad Consent in der Praxis: erste Erfahrungen, aktuelle Entwicklungen und kritische Betrachtungen [Nationally standardized broad consent in practice: initial experiences, current developments, and critical assessment]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2024 Jun;67(6):637-47. DOI: 10.1007/s00103-024-03878-6

15. Bundesministerium für Gesundheit (BMG). Gesundheitsdatennutzungsgesetz (GDNG). 2023 [cited 2025 Apr 30]. Available from: https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/detail/gesundheitsdatennutzungsgesetz.html

16. European Parliament. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 (Text with EEA relevance). 2025 Feb 11 [cited 2025 Apr 30]. Available from: https://data.europa.eu/eli/reg/2025/327/oj/eng

17. Karagiannis S, Ntantogian C, Magkos E, Tsohou A, Ribeiro LL. Mastering data privacy: leveraging K-anonymity for robust health data sharing. Int J Inf Secur. 2024;23:2189-201. DOI: 10.1007/s10207-024-00838-8

18. Bray T, editor. The JavaScript Object Notation (JSON) Data Interchange Format, STD 90, RFC 8259. RFC Editor; 2017. DOI: 10.17487/RFC8259

19. JSON Schema. [cited 2025 Apr 17]. Available from: https://json-schema.org/

20. creative commons. About CC Licenses. [cited 2025 Apr 17]. Available from: https://creativecommons.org/

21. Buhl M. Interpretable Clinical Decision Support System for Audiology Based on Predicted Common Audiological Functional Parameters (CAFPAs). Diagnostics (Basel). 2022 Feb 11;12(2):463. DOI: 10.3390/diagnostics12020463

22. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. arXiv. 2017. p. 1273-82. DOI: 10.48550/arXiv.1602.05629

23. Pfitzner B, Steckhan N, Arnrich B. Federated learning in a medical context: a systematic literature review. ACM Transactions on Internet Technology (TOIT). 2021;21(2):1-31. DOI: 10.1145/3412357

24. Deist TM, Dankers FJWM, Ojha P, Scott Marshall M, Janssen T, Faivre-Finn C, Masciocchi C, Valentini V, Wang J, Chen J, Zhang Z, Spezi E, Button M, Jan Nuyttens J, Vernhout R, van Soest J, Jochems A, Monshouwer R, Bussink J, Price G, Lambin P, Dekker A. Distributed learning on 20 000+ lung cancer patients - The Personal Health Train. Radiother Oncol. 2020 Mar;144:189-200. DOI: 10.1016/j.radonc.2019.11.019

25. Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, Somani S, Paranjpe I, De Freitas JK, Wanyan T, Johnson KW, Bicak M, Klang E, Kwon YJ, Costa A, Zhao S, Miotto R, Charney AW, Böttinger E, Fayad ZA, Nadkarni GN, Wang F, Glicksberg BS. Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. JMIR Med Inform. 2021 Jan 27;9(1):e24207. DOI: 10.2196/24207

26. Saak S, Huelsmeier D, Kollmeier B, Buhl M. A flexible data-driven audiological patient stratification method for deriving auditory profiles. Front Neurol. 2022 Sep 15;13:959582. DOI: 10.3389/fneur.2022.959582

27. Saak S, Oetting D, Kollmeier B, Buhl M. Integrating Audiological Datasets via Federated Merging of Auditory Profiles. Trends Hear. 2025 Jan-Dec;29:23312165251349617. DOI: 10.1177/23312165251349617

28. Smola AJ, Schölkopf B. Learning with kernels. Vol. 4. Berlin: GMD-Forschungszentrum Informationstechnik; 1998.

29. Tanabe H, Bao Ho T, Nguyen CH, Kawasaki S. Simple but effective methods for combining kernels in computational biology. In: IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies; 2008 Jul 13-17; Ho Chi Minh City, Vietnam. IEEE; 2008. p. 71-8. DOI: 10.1109/RIVF.2008.4586335

30. Berg D. Forschungsdatenmanagement in der Audiologie: Status – Entwicklungen – Spannungsfelder. In: Deutsche Gesellschaft für Audiologie e. V.; ADANO, editors. 27. Jahrestagung der Deutschen Gesellschaft für Audiologie und Arbeitstagung der Arbeitsgemeinschaft Deutschsprachiger Audiologen, Neurologen und Otologen. Göttingen, 19.-21.03.2025. Düsseldorf: German Medical Science Publishing House; 2025. Doc012. DOI: 10.3205/25dga012

Corresponding author:

Daniel Berg
Carl von Ossietzky University Oldenburg, Ammerländer Heerstr. 114–118, 26129 Oldenburg, Germany
daniel.berg@uni-oldenburg.de

Please cite as
Berg D, Buhl M, Krefting D. Research data management – chances, challenges and solutions for audiology. GMS Z Audiol (Audiol Acoust). 2025;7:Doc11.
DOI: 10.3205/zaud000074, URN: urn:nbn:de:0183-zaud0000744

This article is freely available from
https://doi.org/10.3205/zaud000074

Published: *2025-11-17*