OLDENBURG HEARING HEALTH RECORD (OHHR)

Sumbul Jafri, Daniel Berg, Mareike Buhl, Matthias Vormann, Samira Saak, Kirsten C. Wagener, Christiane M. Thiel, Andrea Hildebrandt and Birger Kollmeier

DOI: 10.5281/zenodo.14177902

Why Open Data Matters in Health Research?

Imagine if hearing care was as personalized as your music recommendations

To get there, we need:

- Rich, multidimensional health data
- Collaboration across disciplines
- Responsible and open data sharing

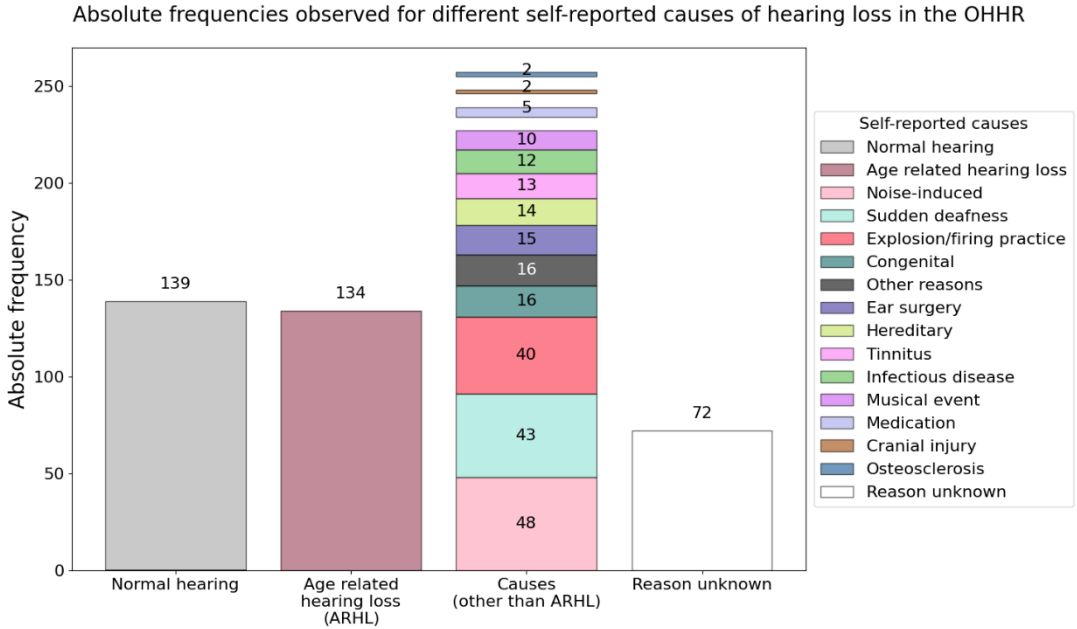
Oldenburg Hearing Health Record (OHHR)

A large dataset of 581 adults with and without hearing loss

(18–86 years; Mean age = 67.31, SD = 11.93; 44% female)

- Rich with audiological, cognitive, health, lifestyle and demographic data
- Fully anonymized
- Interoperable and well-documented
- Openly available on Zenodo under CC BY 4.0 license

DATA



Data Type	Details
Home Questionnaire[2,3,4] (self-report)	Quality of Life: Hearing anamnesis, SF-12[5], chronic conditions, hearing aid usage and impact on daily life, Technological readiness[6] and media consumption habits
Clinical Interview (conducted by an expert)	Hearing health history: Hearing aid or implant use, familial hearing loss, ear infections, socio-demographic information
Pure Tone Audiometry	Hearing threshold levels (air & bone conduction) and Uncomfortable Loudness levels across 125–8000 Hz for left and right ear
Adaptive Categorical Loudness Scaling[7]	Measurements recorded for 1500 and 4000 Hz narrow-band noise stimuli for the left and right ear
Speech-in-Noise Measures	Göttingen Sentence Test[8] and Digit Triplet Test[9]
Cognitive Measures	DemTect[10](dementia detection), Vocabulary size test (Wortschatz[11]; proxy for crystallized intelligence)

ANONYMIZATION

Protecting Participant Privacy:

Data Collection

- **Informed Consent**

  → Voluntary, documented participation with the right to withdraw
- **Pseudonymization**

  → Direct identifiers are removed and replaced with coded ID

  → Re-identification is still possible via secure linkage
- **Anonymization**

  → All identifiers are removed (name/contact details)

ANONYMIZATION

Data Curation

- **Pseudonymized data with contact option:**

  → Recontacted to get consent for full data release

  → After consent, linkage keys were removed

- **Already anonymized data (no recontact possible)**

  → No direct identifiers remained

  → Applied **k-anonymity** (k = 4) to minimize re-ID risk

  → Quasi-identifiers (e.g. education/occupation) put into broader categories

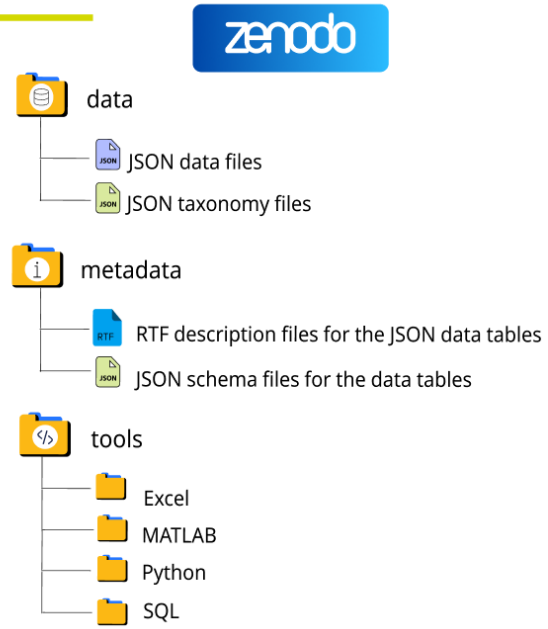
  → Approved by data protection officer

DATA RECORDS

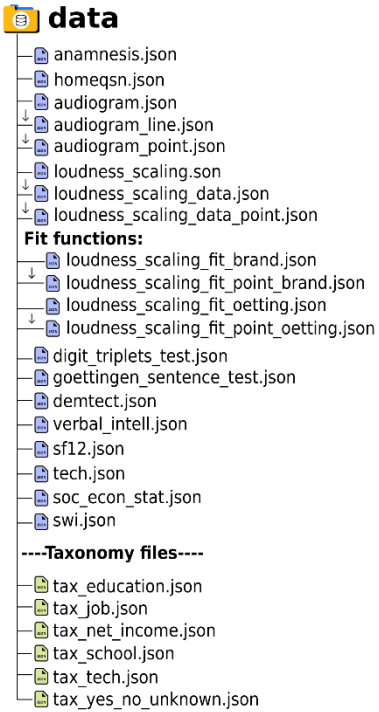
Data: Machine-readable JSON files on “Hearing4all” Zenodo community ensure easy discovery and open access

Metadata: Rich metadata, standardized taxonomies, and clear licensing enable interoperability and reusability

Tools: Custom scripts for various platforms allow flexible adaptation for future research



STRUCTURE



- Modular JSON tables, linked by IDs
- Taxonomy files for consistent, machine-readable categories

```
[
  {
    "id": 100000,
    "value": 0,
    "name": "no"
  },
  {
    "id": 100001,
    "value": 1,
    "name": "yes"
  },
  {
    "id": 100002,
    "value": 2,
    "name": "unknown"
  }
]
```

STRUCTURE - EXAMPLE

audiogram.json

```
[
  {
    "clientid": 100000,
    "audiogramid": 100000,
    "date": "2014-04-15"
  },
```

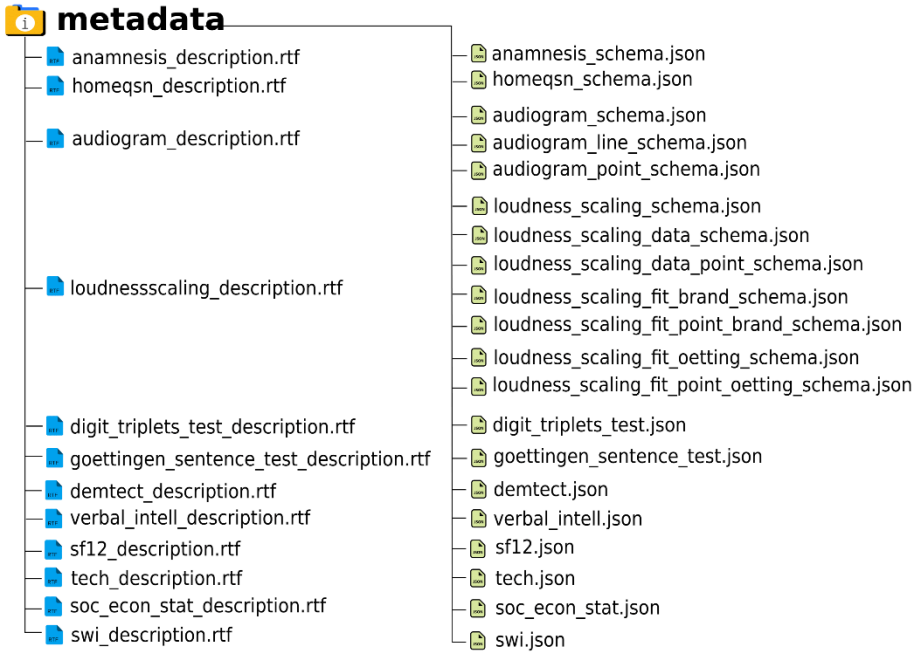
audiogram_line.json

```
[
  {
    "audiogramlineid": 100000,
    "audiogramid": 100000,
    "transducertype": "ac",
    "transducername": "hda200"
    "reference": "dB HL",
    "side": "right",
    "signaltype": "sine",
    "masking": false,
    "type": "htl"
  },
  {
    "audiogramlineid": 100001,
    "audiogramid": 100000,
    "transducertype": "ac",
    "transducername": "hda200"
    "reference": "dB HL",
    "side": "left",
    "signaltype": "sine",
    "masking": false,
```

audiogram_point.json

```
[
  {
    "audiogramlineid": 100000,
    "frequency": 125,
    "level": 35,
    "status": "Normal",
    "maskinglevel": null
  },
  {
    "audiogramlineid": 100000,
    "frequency": 250,
    "level": 25,
    "status": "Normal",
    "maskinglevel": null
  },
  {
    "audiogramlineid": 100000
```

METADATA



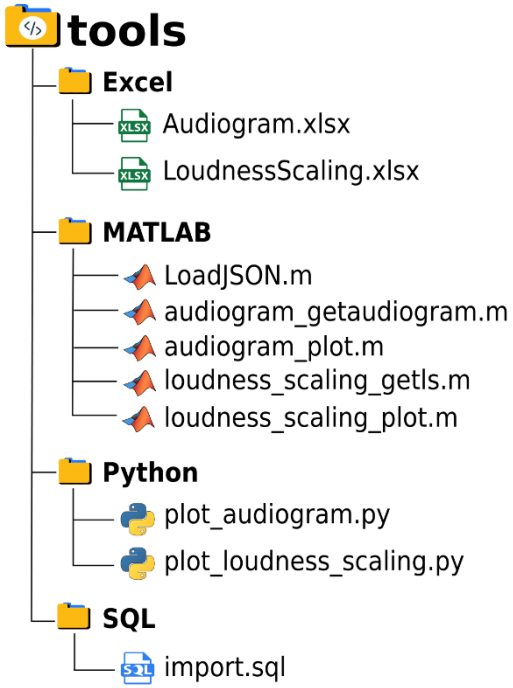
- RTF docs describe collection & context
- Schema files define JSON table structure

METADATA - EXAMPLE

verbal_intel_schema.json

```
{
  "$schema":"https://json-schema.org/draft/2020-12/schema",
  "$id":".",
  "title":"VerbalIntelligence",
  "description":"Describes one verbal intelligence test for one client",
  "type":"array",
  "items":{
    "type":"object",
    "properties":{
      "id":{
        "description":"Unique identifier of the test",
        "type":"integer"
      },
      "clientid":{
        "description":"Identifier of the client",
        "type":"integer"
      },
      "date":{
        "description":"Measurement date of test",
        "type":"string",
        "format":"date"
      },
      "score":{
```

TOOLS



- Example scripts in Excel, MATLAB, Python, and SQL
- Help with importing and plotting audiological data
- Lower the barrier for users with different backgrounds

PUBLICATION

- Preprint on *medrxiv* – https://doi.org/10.1101/2025.03.30.25324761
- Manuscript under consideration at Scientific Data
- Dataset with documentation is freely available on Zenodo

RECOMMENDATIONS

- Get broad consent
- Involve DPOs & domain experts early
- Refer to legal frameworks (GDPR, GDNG, European Data Governance Act)
- Use FAIR formats & open licenses
- Structure data relationally
- Provide tools for reuse
- Publish as openly as possible

REFERENCES & RESOURCES

1. Jafri, S., Berg, D., Buhl, M., Vormann, M., Saak, S., Wagener, K. C., Thiel, C., Hildebrandt, A., & Kollmeier, B. (2025). OHHR – The Oldenburg Hearing Health Record [Dataset] (1.2.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15727482

2. Berg, D., Jafri, S., & Hildebrandt, A. (2025, February 14). Publishing medical data: The Oldenburg Hearing Health Repository as an example on data re-organization, anonymization and tooling. Zenodo. https://doi.org/10.5281/zenodo.14855523

3. JSON Schema, „JSON Schema“, https://json-schema.org/

4. How-to-Fair.dk, „File formats“, https://howtofair.dk/how-to-fair/file-formats/

5. DataCite, „DataCite Metadata Schema“, http://schema.datacite.org/

6. NFDI4Health, „NFDI4Health Metadata Schema“, https://www.nfdi4health.de/en/service/data-harmonisation-and-publication.html

7. re3data, „Registry of Research Data Repositories“, https://www.re3data.org/

8. creative commons, „About CC Licenses“, https://creativecommons.org/

Thanks for listening