



14th IC BEN Congress on Noise as a Public Health Problem



Launch of IC BEN's brand new survey data archive ISAR

Mark Brink¹, Thu Lan Nguyen², Dirk Schreckenberg³, Jördis Wothge⁴

¹ Swiss Federal Office for the Environment, CH-3003 Bern, Switzerland

² Shimane University, Shimane 690-8504, Japan

³ ZEUS GmbH, D-58095 Hagen, Germany

⁴ German Environment Agency, D-06844 Dessau-Rosslau, Germany

Corresponding author's e-mail address: mark.brink@bafu.admin.ch

ABSTRACT

The IC BEN's officers and Team 6 (Community response to noise and annoyance) chairs have started the project of setting up an international archive of original survey response data from socio-acoustic noise surveys, named ISAR (short for IC BEN Socio-Acoustic Survey Archive). The archive caters to noise effects researchers around the globe and will enable pooled analyses, establishment of generalized exposure-response relationships, trend analyses both spatially and temporally, as well as cross-cultural comparisons of self-reported noise reactions, such as annoyance or self-reported sleep disturbance. Thus the data in the archive are freely accessible. All data in the ISAR archive are fully anonymous in order to meet the pertinent data protection regulations. The ISAR archive explicitly collects only raw survey response data, rather than data that have been aggregated on a higher level. We are constantly seeking to extend the archive with new study data and thus, every new contribution is highly welcome. In this IC BEN talk, we would like to publicize and promote the archive and encourage our colleagues to share their survey data with the archive, and of course, later to use the data for exciting new research on the biological effects of noise.

Keywords: IC BEN, Survey, Archive, Annoyance, Pooled Analysis

INTRODUCTION

The International Committee on Biological Effects of Noise (IC BEN) starts off an open archive of original exposure and response data from so called socio-acoustic noise studies (or surveys), called ISAR (IC BEN Socio-Acoustic Survey Data Archive). The ISAR archive in its current state primarily focuses on noise annoyance surveys and caters to noise effects researchers around

the globe. The archive will enable pooled analyses, establishment of generalized exposure-response relationships, trend analyses both spatially and temporally, as well as cross-cultural comparisons of noise reactions. We believe that scientists in our field should rather share their data once their original scientific exploitation is completed, than just keeping them for themselves indefinitely. With this initiative, IC BEN is also committed to support the current trend in science towards the direction of "open data", and thus simply, good scientific practice. Bluntly, overcoming monopolies on data sets is the order of the day in our view. By promoting the widest possible access to the data within our community, the already scarce resources mostly financed by taxes are thus better used and for the benefit of noise effects research as a whole. Yes, even more and more funders are jumping on this bandwagon and demanding that publicly funded data collections be made available to the public in turn, sooner or later. Thus, the archive will be openly accessible to everyone and tries to adopt a FAIR (Findable, Accessible, Interoperable, Reusable) data culture as far as possible. The archive will be published under a Creative Commons license and – as currently planned – on the ZENODO platform (zenodo.org), a general-purpose open repository that is free and run by CERN (European Organization for Nuclear Research). The first version of the archive is planned to be released in June 2023, on the occasion of IC BEN's 14th congress and its 50th anniversary. Guidance as pertaining to how the archive can be used and how one can contribute to it, will be provided on the official IC BEN website (www.icben.org/isar)

Note: The setting up of the ISAR archive is a currently ongoing process. This short paper reflects the current state of the project. The final form and the modalities of the archive is subject to change.

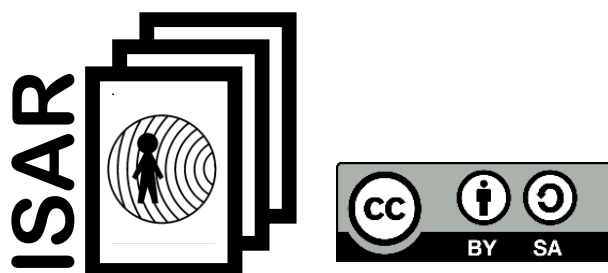


Figure 1: ISAR logo and Creative Commons CC BY SA icon

DATA CURATION AND ACCESS TO THE ARCHIVE

IC BEN is an organization of noise effects researchers and as such is interested in serving its community with an up-to-date high-quality archive of individual response data from past and future socio-acoustic surveys. To reach this, IC BEN assumes the role as a data curator within the ISAR project. This means that IC BEN (1) requests from peers in the noise effects research community to make their data available for the archive, (2) enforces minimal formatting and documentation requirements of data that are delivered to the archive, (3) organizes the data, (4) documents the archive, and (5) makes the archive publicly available in a convenient form (i.e. as an easily accessible online repository). The goal of this process is that the value of the archive is maintained in each new version over time, and that the data in the archive remain available for reuse within the scientific community. The archive shall basically be curated (managed) by IC BEN representatives (officers and team chairs of Team 6), but any other IC BEN-affiliated person can also act as data curator of the ISAR archive. Organizational matters are currently worked out.

Once the data curation activities for a new version of the archive are finished, the archive will be published (physically stored) on the ZENODO platform and will receive an unequivocal and persistent digital object identifier (DOI). That means: For each new version of the archive (e.g. after another survey entered the archive), a new DOI is minted, which makes each version of the archive easily citable. For maintaining full transparency, it is planned that all versions of the archive ever published will remain accessible and individually citable through their individual DOIs. Access to the archived data on ZENODO will be free for everyone, and everyone will be able to use the archived data for scientific purposes.

DATA PROTECTION ISSUES AND LICENSING

In the wake of the first data collections for the archive, we consulted several legal experts who confirmed that there shouldn't be any fundamental problems with potential data protection violations if the data in the archive are fully anonymous. This means that the data must not contain any personally identifiable information and that the combination of variables must not allow re-identification of a person. A person is considered unidentifiable or no longer identifiable if it would require a significant effort to identify them. It is also important to note that anonymized data do not fall under the regulations of the GDPR (General Data Protection Regulation of the EU), which means that far less restrictions regarding the public dissemination of such data apply. The ISAR archive is planned to be published under a CC Attribution Share-Alike License (CC BY (SA)). This license requires the data user (the licensee) to cite or credit the data source (i.e. the ICBEN/ISAR archive) in all works that are based on data in the archive, in the manner specified by ICBEN. Furthermore, data users may distribute derivative works (e.g. overview tables derived from data in the archive) only under a license identical to ("not more restrictive than") the CC BY SA license that governs ISAR. Yet, there is an ongoing discussion regarding the "SA" feature, and we may decide to strip it. ISAR data can also be used commercially, as long as the commercial activity pursues a basically scientific goal. This means, for example, that a private environmental consulting firm can use the data in the archive to produce a review on noise effects commissioned and paid by, e.g., a government agency.

TERMS AND CONDITIONS FOR DATA PROVIDERS AND DATA USERS

Aside from ICBEN as data curator, two other roles are connected to the archive: data providers and data users:

As the ISAR archive is an open research data repository, the **data providers** are responsible to ensure that the content they make available for the archive is suitable for open dissemination and that it complies with applicable laws, including privacy, data protection, and intellectual property rights. Such data have been collected by the data provider or its institution in accordance with the local data protection regulations and have been anonymized to an appropriate degree. Data providers retain all their rights on their data and ICBEN does not acquire ownership of data or copyright. Of course, the relinquishing of data into the archive requires a certain effort on the side of the data providers as they have to select the suitable variables, do some formatting on data tables etc. and provide a minimal amount of documentation. In return, data providers can expect that studies in the ISAR archive will more likely be cited in works that are based on the archive. Quite certainly, the larger the archive will grow, the higher its scientific utility will be, again boosting the visibility of original studies in the archive. The people at ICBEN who are responsible for the archive are, hopefully, happy to provide as much support as necessary for the transfer of the data into the archive.

Whenever a data provider releases their study data into the archive, a contract is set up that must be signed by an ICBEN representative as well as by the data provider. Data providers must explicitly agree that their delivered data will, as part of the archive, be published online and that the data will be made available under a CC BY (SA) license.

Data users are persons or institutions that access the data in the archive for scientific purposes, e.g., for carrying out pooled analyses across all or a selection of data in the archive in the context of overview articles, reviews and similar undertakings. As the ISAR archive adopts an open data philosophy and should be openly available to everyone, there are no specific restrictions ICBEN can impose on data users. This means that anyone can download the archive and analyze its data without consent from ICBEN. For that reason, no formal (mutually signed) contract is established between ICBEN and data users.

The detailed terms and conditions for both data providers and data users are currently being worked out.

TYPE OF DATA THE ARCHIVE SEEKS

The ISAR archive explicitly aims at collecting original (but anonymized, see above) raw data from noise surveys of all modes (paper-and-pencil, telephone, personal, online, or hybrid) on peoples self-reported reactions to noise. This is the maximum level of detail an empirical dataset can have (as opposed to data that have been aggregated on a higher level, e.g. the percentage highly annoyed within an exposure level category). Only raw data, i.e. one record per respondent of a survey, allows to pool original data from different studies in a 1:1 fashion.

The main focus of the archive is on annoyance and self-reported sleep disturbance, but basically all other variables that have been collected in a survey are highly welcome and will be included in the archive as well. However, it is (almost) fully up to the data providers to decide which variables they want to share with the archive, with few exceptions: In order to allow for the derivation of exposure-response relationships, data about noise exposure are a necessity. Thus, the absolute minimal set of (mandatory) variables to provide so that a study can be included in the archive is: At least one noise source, one noise exposure metric, and one noise reaction/response variable. Further important variables that are highly desired to be reported are year of birth (or age or age category at the time of questionnaire completion), gender, noise sensitivity, survey year, and geographic area (country, city). Each data delivery should be accompanied by a description of the variables, type of response scale used, values assigned to responses, and the original wording of the question.

It is important to note that the ISAR archive solicits entries not only from recent and future new studies, but explicitly also welcomes original data from studies that date back several years or decades. Such surveys are a highly valuable source when it comes to analyze trends in, e.g. noise annoyance over time, a subject that fostered particular debate in recent years.

NEXT STEPS AND FUTURE PROSPECTS

As long as there will be enough people at ICBEN that are committed to maintaining the archive, ICBEN will in the future constantly seek to extend the archive with new study data. We have already started to collect data for the archive from several peers in the noise effects research community, totaling almost 40'000 individual responses. The main activity to follow in the short-term will be the finalization of important documents governing ISAR (General Terms and Conditions, Data provider contract) and their endorsement by the ICBEN executive committee, the further extension of the archive, and hopefully the publication of its first release around mid-

year 2023. In the long-term, we believe that pooled analyses, such as will be possible with the consolidated ISAR data, will in the future more thoroughly inform general noise policies on an international level than was possible in the past. And one day, the ISAR archive could provide a valuable source for e.g. an eventual update of the WHO Environmental Noise Guidelines.

Acknowledgements

We would like to thank our colleague Matthias Lochmann for fruitful discussions and useful advice regarding the physical storage and licensing options of the archive. We would also like to thank other colleagues from ICBEN for their comments and support, and the legal experts we contacted for their opinions on legal issues.