Reproducible Science and Research Integrity Policy

of the
Learning, Educational Achievement, and Life Course Development
Graduate School & Research Network (LEAD)
and the
Hector Research Institute of Education Sciences and Psychology (HIB)

June 2017

Contents

Goals

Good science

LEAD and HIB are committed to secure and advance good science and therefore strive to make research results robust, reliable and useful. Systematic documentation, sustainable conservation and structured dissemination are key components in the pursuit of a truly cumulative and integrative body of knowledge in a world of ever more rapid data creation and availability.

Efficient use of data

Empirical data serves the goals of answering substantial and important questions, or enabling researchers to derive and formulate research questions. But their collection requires spending of considerable resources (time, effort, but also financial in nature) and the participation of many volunteers that may be inconvenienced or even strained to different degrees by their participation. In order to use data in an efficient and sustainable way that minimizes the strain on participants and resources, data must be used efficiently.

Multiple use

Data should be used to answer different research questions from different (intra- or inter-disciplinary) perspectives, using different methodological tactics, strategies, and tools, and focusing on different aspects. In order for data to be used in multiple investigations, it must be well documented and available in structured formats that ensure interoperability at a high level of feasibility.

Such other uses for the researcher originally conducting a study, close and remote collaborators, as well as the scientific community in general include quite prominently (but are not limited to) meta-analyses, re-analysis for different research questions and hypotheses, re-analysis using different novel methodological tools and strategies that were not yet available at the time of the first original data analysis, exemplary use for teaching and training purposes.

Sustainability

Data can not only be used concurrently by researchers from different disciplines or with different perspectives, but it may also be re-used in the future, in ways that may or may not be readily conceivable at present. First tests with newly developed methodological tools, first tests of truly novel and hitherto not conceived hypotheses, usage of variables that have not received special attention so far are examples of such future usefulness potential. Since future developments cannot – by definition – be foreseen at a present point in time, conservation of data sets for future use is indispensable if this future potential is to be maintained. Therefore sustainability of documentation and archival of research data is a priority. Such sustainability is achieved through the use of simple, robust, open formats and structures that ensure concurrent and temporal interoperability.

Protection

Recent developments in the social sciences have led to cases of accusations, doubts and recriminations of individual research practices. In order to assure researchers and protect them from a lengthy and costly process of defending their research practice against external concerns of questionable practices, LEAD and HIB are

committed to a research data documentation and archival that allow their members to quickly, relatively effortlessly and confidently refute such illegitimate accusations. Therefore, the comprehensive, systematic, and well documented archival of research data is emphasized as an instrument to protect researchers and give them safety to go about their core business of doing research instead of worrying about having to struggle with unfounded accusations.

Documentation & Accountability

LEAD and HIB welcome the recent and increasing trend of funding agencies to postulate the documentation and accountability for resources that are made available from public or private funds in the course of research grant processes. In order to meet the requirements that are imposed onto grant application processes and administration, the two institutions seek to offer the logistics and support necessary for individual researchers to fulfill the requirements.

<div align="center">**Means**</div>

<div align="center">*Research data archival*</div>

Data and analysis documentation

All original and raw data collected in the course of research projects at LEAD and HIB are to be stored in a comprehensive, robust, and accessible interoperable way. Similarly, any transformation and conversion, selection, and analysis steps in the course of data processing should be documented. Data and data processing documentation is to proceed in a way that would in principle enable an informed third party to redraw and reproduce the same results from the same data as the original analysis. This does not imply that all research data will necessarily be made publically available. Rather, any archived research data should be stored in a way that is generically intelligible to

- the original researcher him- or herself after focusing on a different research project for a few months while waiting for a decision about a manuscript submitted for publication,
- a collaborator within the same institution or work group,
- external collaborators,
- future members of the institutions or the scientific community in general, but also
- editors and reviewers in cases where inclusion of research data is required when submitting a manuscript.

Materials

To in principle enable an informed third party to comprehend, review, reconstruct or even replicate a particular study, it is vital to document all materials used in a study process. Again, such use by a third party may not necessarily be undertaken or planned, but research data should be archived so that it will be immediately possible should this change. Materials include, stimuli, original questionnaire formats, item lists, programmed procedures (e.g., a presentation sequence in an experimental software), details about procedures pertinent to how and in which temporal or logical sequence manipulations, interventions, and measurements were conducted and administered, collection or selection algorithms for incidental and archival data (such as computer logs, large bodies of verbal or textual data, etc.).

Ethics documentation

Every study conducted at LEAD and HIB must undergo an ethical approval process with one of the official ethics commissions (IRB) at the Eberhard Karls Universität Tübingen, the University Hospital Tübingen, or the Leibniz-Institut für Wissensmedien Tübingen (additional approval processes may be required, e.g. in the case of data collection in schools). This approval guarantees the conformity of research procedures with appropriate guidelines and rules and relieves researchers of uncertainty and worries regarding the ethical innocuousness of their research. For all studies conducted under the auspices of LEAD or HIB, the process leading to the ethical approval by an appropriate institution should be fully documented with the research data, independently of where the research data are ultimately archived.

Location

Research data is preferably archived in the Research Data Portal of Eberhard Karls Universität Tübingen (https://fdat.escience.uni-tuebingen.de/portal/). Details about the structure of archived research data and ways to commit research data into the repository will be detailed separately. In case they have a systematic research data archival policy in place, institutions that are involved or associated with LEAD may use different repositories. In any case, the research data from all studies conducted under the auspices of LEAD or HIB will be systematically archived according to a systematic research data archival policy in any one pertinent location.

Data sharing

LEAD and HIB are committed to a cumulative science process that advances through efficient use of research data. Therefore, overall a maximum of research data should be shared among a scientific community at least, but ideally without restriction, provided that the participants who have provided the data and their interests as well as general legal interests are appropriately protected. Exceptions to this rule are specific and circumscribed.

Location

Research data can be widely published or shared with a number of specific individuals using the Research Data Portal of Eberhard Karls Universität Tübingen. This repository ensures appropriate persistence, reliable accessibility, security and dissemination of meta data as well as the research data proper, should it be made openly available without restrictions. Other repositories offer data sharing infrastructure and may be used instead. However, care should be taken to avoid unnecessary duplication and ensure consistency of multiple copies at different institutions and repositories with data originally archived at Eberhard Karls Universität Tübingen.

Preregistration

LEAD and HIB emphatically encourage the pre-registration of experimental as well as non-experimental studies. In the former case, theorizing, derivation of hypotheses and candidate models, data collection and analysis methods are documented and logged before data collection begins. In the latter case data collection may often have been previously completed, in which case theorizing, derivation of hypotheses and candidate models, and analysis methods can be fully logged before analysis begins.

Contrary to a common misunderstanding, pre-registration of studies does not stifle or devalue exploratory analyses. Pre-registration simply marshals analysis results into those that were genuinely predicted before the analysis and those that were not, but in no way strips researchers of the freedom they enjoy in creatively exploring data without clear hypotheses. On the contrary, pre-registration explicitly distinguishes between hypothesis-testing and exploratory investigation and therefore characterizes both types of analysis as on par and of equal importance. Without such a distinction, exploratory research and analysis are often incorporated into a hypothesis testing framework up to the degree where it vanishes as legitimate and original exploratory research (e.g., by basic and very normal human hindsight and faulty memory processes).

Location

Ideally, preregistration should be filed with an external institution or website that establishes a clear and unambiguous time line of registration events. A number of locations afford preregistration infrastructure, such as the Open Science Framework OSF (www.osf.io) or AsPredicted (www.aspredicted.org).

Effective and efficient research designs

A chief goal of scientific research is to obtain answers to questions that can be relied and acted upon with a measure of confidence, regardless of what the specific answer may be one obtains through empirical investigation. In order to fulfill this goal, individual research projects must be designed in a way that will assure researchers that the answer that will obtained from the subsequent data analysis is a valuable piece of information regardless of its specific content, rather than leave them hoping for a particular result (and possibly inadequately color their judgment in specific analysis decisions, thus possibly inadvertently biasing the results in favor of a specific result). In this vein, the simultaneous systematic consideration and control of Type I and Type II error rates, minimally interesting effect sizes and appropriate sample sizes (not to small, but also not wastefully large) is to be given specific attention. But also hypotheses and research questions should be carefully and diligently derived and formulated and adequately cast into the design so as to ensure that the answers to be contained optimally correspond to the questions asked. This process of deriving study designs from research questions and hypotheses should be systematically and comprehensively documented, if only so that researchers' original theorizing, ideas and thinking cannot be lost by degrading memory.

Protection of research participants

Human individuals who volunteer to participate in the research conducted at LEAD and HIB are treated with the utmost respect and care. This means that researchers at the institutions take all possible measures to ensure …

- … protection of participants' privacy and anonymity as well as confidentiality of the data they provide to the highest degree possible, e.g.:
    - Separation of individuating and de-anonymizing information from other data at the earliest possible point in time with double checks before archival of data to ascertain that archived data contains no violations of anonymity
    - In case absolute anonymity is temporarily constrained, procedures and data protection management is systematically planned, implemented and well-documented to minimize temporary risks and threats to anonymity (e.g., by replacing individuating information by non-descriptive codes at the earliest opportunity and safe storage and management of information mapping individuating information to said codes by a trustworthy external agent not otherwise involved in the research. Such mapping information is to be kept in one exclusive single place and as one exclusive single copy
    - No unnecessary duplication, copying and sharing of data, especially privacy sensitive portions thereof
- … transparency,

- o Participants are thoroughly and carefully informed about the research they participate in, including stages of the research after data collection proper (analysis, evaluation, publication, archival, sharing and potential re-use of the collected data)
  - o Informed consent to participation of all research participants (and potentially their legal guardians) is obtained and documented as required
- … and protection of participants
  - o from any strain, stress, psychological or physical discomfort or even pain that they did not consent to in written form after having been thoroughly and openly informed about it
  - o Independently of information and informed consent by participants, LEAD and HIB are committed to absolute abstinence from inflicting harm on participants that is not justified and deemed necessary based on specific and strong scientific argument.

Definitions

Research data

Research data are all pieces of information that pertain to a particular empirical study and enable an in principle uninvolved, but informed third party to fully understand the study, come to similar conclusions applying the documented procedures and in principle replicate the study. They comprise, for example,

- used materials
- design information (including assignment/randomization protocols, description of the pertinent details regarding data collection contexts and settings)
- original data (documentation of the data from the first time observations become manifested in the form of responses or numbers)
- raw data (a full transcript of all observations into numbers and or text strings)
- data transformation documentation (any and all filters, results of scaling and index calculation procedures, recoding into true correct or incorrect answers)
- documentation of analyses, both pertaining to hypothesis tests and exploratory
- ethics approval documentation

Meta data

Any information that describes the content and structure of research data is meta data. Meta data allows to search, find, index, survey, use and re-use original research data and should be stored in a set structure and format to ensure sustainability, convertibility and interoperability. The usefulness of research data declines rapidly with increasing degrees of idiosyncrasy in the contents and structure of meta data. The ultimate goal of meta data storage is to allow for the reproduction of research in ideally all steps of the process from the first inception of a study through any final published results by an informed third party, using research data and associated meta data, and without reliance on the presence of a particular individual involved in the originally conducted study. This goal may not be completely attainable in practice, but efforts should always be directed toward the goal rather than away from it.

Interoperability

"Interoperability is a characteristic of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, in either implementation or access, without any restrictions." (<https://en.wikipedia.org/wiki/Interoperability>). The higher the ability of users to re-use information using different technologies, software packages and systems, the higher is the interoperability of research data. Proprietary and closed file types and fragmented idiosyncratic structures and formats are detrimental to interoperability.