DATA MANAGEMENT PRACTICE

5.1 DATA MANAGEMENT PRACTICE

Background

Conducting research responsibly is more than avoiding intentional fabrication or falsification of data. This is because data provide the factual basis for scientific work and the value of research depends directly on integrity in the collection, use, and sharing of data.

Policies or guidelines which address the ownership of research materials and data, their storage, their retention beyond the cessation of the project and appropriate access to them by the research community is required. It is important to retain the research data because it may be all that remains of the research work at the end of the project. While it may not be practical to keep all the primary source of materials (e.g. biological materials, test results, questionnaires or recordings), durable records derived from them (e.g. assays, test results, laboratory and field notes and transcripts) must be retained and remain accessible. The researcher must then decide which data and materials should be retained. However, the retaining of such materials could be determined by institution's policies and guidelines, law and or funding agencies. (Do confer with your institution's guidelines and or policies.)

The main aim of retaining sufficient research data and materials is to justify the research outcomes and to defend them if they are challenged. Therefore, the potential value of the data and materials for future research should also be considered, especially where the research would be difficult or impossible to repeat.

What is research data?

Data is defined as a collection of facts, measurements, or observations used to make inferences about the world we live in. However, data in research can range from material created in a wet laboratory, such as an electrophoresis gel or a DNA sequence, to that obtained in social-science research, such as a completed questionnaire, videotapes, and photographs. Research data can include microscope slides, cell lines, climate patterns, soil samples, astronomical measurements, and spectrographic analyses. In addition, custom software or hardware and specialized methods can be considered as data too.

Who owns the research data?

Identifying ownership of research data and primary materials

Although graduate students, postdoctoral fellows, or even some faculty in academia performing research may believe that they own the data collected, they are wrong. Research institutions should have policies or guidelines on the ownership of the research data and materials acquired by staff during research. The ownership may be influenced by the possible funding arrangements for the project. However, as a general rule of thumb, all materials and data retained at the end of a project are the property of the institution that hosted the project. That would be the most satisfactory arrangement.

For example, if the host institution is Institution ABC, than all materials and data retained at the end of the project is the property of Institution ABC.

The Pl takes responsibility for the collection, recording, storage, retention, and disposal of data. When data is published, the copyright is retained by the Pl, who then assigns it to the publisher of the journal. Data and data books collected by undergraduates, graduates, and postdoctoral fellows on a research project belong to the host institution, and students should not take their data when they leave without making appropriate arrangements. Retaining copies of data is allowed with permission, and although this is not always done, it is certainly good practice.

When study team members leave an institution, they have to negotiate with the institution to keep their grants and data. With industry funded or privately funded research, data can belong to the sponsor, although the right to publish the data may or may not be extended to the investigator.

Do you know?

The Bayh-Dole Act
The Bayh–Dole Act or Patent and Trademark Law Amendments Act is a United States legislation dealing with intellectual property arising from federal government-funded research.

The Bayh-Dole Act of 1980 allowed universities to have control of the intellectual property, such as patents, generated from federally funded research. With a patent in hand, universities could exclusively license the patent to businesses. For the past 25 years or so, many universities, including Columbia University, have benefited from the licensing revenue. Recent inventions, in the form of new drugs and computer technologies, have also helped the public. The law has encouraged new relationships between academic researchers and companies, but critics such as Derek Bok (2003) have charged that Bayh-Dole promotes universities' selling out their interests to industry rather than relying on raising money from tuition and other sources.

5.2 BEST WAYS TO COLLECT DATA

Most institutions may not be equipped with the manpower or necessary skills to provide education and training resources needed to formally instruct trainees and junior

researchers in good data-management practices. Therefore, for the scientific enterprise to be productive in the long run there should be positive and comprehensive mentoring of students in data management.

In essence, there is no one way to keep data, but data should explain why research was done, how it was done, where primary data are kept, what happened and didn't happen, an interpretation of the data, and what's next. Data should allow another researcher the ability to repeat the experiment. The data also should be kept in a way that is easy to understand. Legally, federal sponsors of research have the right to audit data and examine records that are relevant to a grant. Data can also be important commercially, in new drug applications to the FDA and for patents on new technologies.

The practice of keeping research notebooks, paper vs. electronic

While many experts recommend collecting data in bound and paper-based notebooks with numbered pages on which the date and time of research can be clearly enumerated, many researchers employ a mixture of electronic and paper based approaches.

While both types of data can be manipulated by someone deciding to engage in misconduct, checks and balances in both can make it harder to do so. The British Medical Research Council advice on good record-keeping states that data should be stored in such a way that it permits a complete retrospective audit, and that it is monitored regularly to ensure completeness and accuracy. The following is a best-practice summary for good record-keeping:

- Raw data should be recorded and retained either in indexed laboratory notebooks with permanent binding and numbered pages or in a dedicated electronic notebook.

- Recording should be done as soon as possible after data are collected and specific note should be made as to whether it represents the date of the recording or the date of collection, if the two are not the same. Modifications should be clearly identified and dated.

- For paper records, a few pages should be kept at the front of a bound book for tables of contents.

- Writing should be done in permanent ink and legibly.

- Copies of original notebooks should be kept elsewhere for safekeeping.

- A second loose-leaf notebook should be kept for data, such as photographs, machine printouts, questionnaires, chart recordings, and autoradiograms that cannot fit into the primary record book.

- Supervisors should review and sign off on notebooks to signify their completeness and accuracy. Queries should be addressed as soon as possible and changes signed by both. Some data may need to be witnessed by a colleague. (Witnessing of data is especially important in commercial research laboratories.)

- Methodology used in an experiment should be written down or a reference to how an experiment deviated from a standard laboratory technique should be explained.

- Lot numbers should be recorded and special attention should be given to the hazardous-substance use.

- Equipment calibrations need to be recorded.

- Data should be noted directly into notebooks without putting it on scraps of paper or relying on memory beforehand.

- All raw data should be included. Be honest.

- Errors should be identified by crossing out the mistakes without obscuring the initial data.

- Material should be logged chronologically.

- Data interpretation should be carefully written.

- Areas in a notebook left blank intentionally should be indicated.

- Correspondence and note conversations related to experiments should be kept.

- Consent forms should be kept with raw data.

- Electronic records need to be carefully monitored.

- Electronic data should be backed up on a disk with a hard copy; relevant software must be retained to ensure future access, should security of data be an issue.

Authorizations to collect data

Institutional Review Boards and human-subject research

If in doubt, researchers should contact NHG DSRB or their institutional ethics review board on how to go about data collection in their research. The NHG DSRB, like IRBs elsewhere, is instituted to ensure the protection of the rights and welfare of human

subjects participating in research at Institutions as well as all institutions under the oversight of NHG DSRB. The boards have the power to approve, disapprove, or modify research protocols involving human subjects. The DSRB operates under the following guidelines and applicable regulations:

- US Department of Health and Human Services (DHHS).
- Regulations 45 CFR 46 – when the research is funded by US Federal Funds e.g. funded by NIH, NCI etc.
- US Food and Drug Administration (FDA) Regulations 21 CFR 56, 21 CFR 50 – when research is being conducted under Investigational New Drug (IND) Application or Investigational Device Exemptions (IDE) or when the results of research are intended to be submitted to FDA.
- International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Guideline for Good Clinical Practice E6 (R1), Current Step 4 version dated 10 June 1996 – All clinical trials involving medicinal products is subject to ICH GCP.
- Medicines Act and Medicines (Clinical Trials) Regulations 2000 – All clinical trials involving medicinal products are subject to the Medicines Act and Medicines (Clinical Trial) Regulations 2000.

The NHG DSRB monitors studies and may suspend or terminate them if there is a danger to subjects or if a researcher is not complying with appropriate guidelines. The NHG DSRB is responsible for determining whether the benefit of the research is sufficient to counterbalance any risks associated with the project. They monitor the nature of the informed consent given to the research subjects as well as the issue of whether confidentiality is maintained during the study and afterwards. NHG DSRB also ensures that special protection is afforded to vulnerable groups, such as children, pregnant women, fetuses and neonates, the cognitively impaired, and prisoners. They also ensure that selection of subjects is equitable and that participation in a research project is voluntary.

Did you know?

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) was enacted by the United States Congress and signed by President Bill Clinton in 1996.
The HIPPA expanded the confidentiality requirements set forth under 45 CFR 46 with respect to patient medical records and information. Researchers looking at clinical data need to know whether they are doing investigations that they can certify will not result in the disclosure of information about a patient; and whether they can obtain a waiver of authorization from a Privacy Board (a committee that looks at HIPAA issues at a university

and other institutions that don't use the IRB for privacy issues) or need authorization from a patient.

5.3 HOW LONG SHOULD RESEARCH DATA AND MATERIALS BE KEPT?

Institutions should provide facilities for the safe and secure storage of research data and for maintaining records of where research data are stored.

According to the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Guideline for Good Clinical Practice E6 (R1), (ICH GCP), it is recommended that essential documents (e.g. Ethics approval letter) should be retained until at least 2 years after the last approval of a marketing application and until there are no pending or contemplated marketing applications, or at least 2 years have lapsed since the formal discontinuation of clinical development of the investigational product, or 6 years after the completion of the clinical trial. These data should be retained for a longer period of time if required by the applicable regulatory requirements and or by an agreement with a sponsor. It is the sponsor's responsibility to inform the investigator/ institution when these documents no longer need to be retained. The documents may be archived by electronic means, microfilm or other suitable archiving technology. It is recommended that the researcher comply with his/her institution's policies or guidelines on how long research data should be stored and store it accordingly.

Did you know?

Legally, data need to be retained for patent protection and in case there is any misconduct allegations pending based on the data. Beyond the professional and legal obligations, researchers in practice will store data as long as they feel it is necessary. But confidential data has to be stored in such a way that access cannot be available. Audits may be necessary to determine whether data have been stored properly. Some investigators store electronic data with archival resources. Minimum storage period is 3 years for all research, and 6 years for clinical trials.

What are the obligations to share data?

There should be a policy on research data ownership and storage. This policy should:-

- Cover all situations possibly anticipated and potentially arising in research, including situations when researchers move to another or between institutions or when data are held outside the institution. Agreements covering ownership and research data storage should be reviewed whenever there is movement or departure of a research staff.

- Whenever appropriate and possible, research data should be held in that researcher's department or other appropriate institutional repository (NHG DSRB Database), although researchers should be permitted to hold copies of the research data for their own use. Agreements for material held in other locations should be documented.

- In projects that involve multiple institutions, an agreement should be developed at the beginning of the research to cover the storage of research data and materials within each institution and these research data and materials must be stored in a safe and secure storage.

Do you know?

In 2003, the National Institutes of Health (NIH) Data Sharing Policy instituted a new policy on data sharing. The new policy applies to investigator-initiated one-year grants and may have an impact on smaller grants too. The goal of the policy is to expedite the timely release and sharing of final data to enhance the research enterprise. Release of data can be complex. Intellectual property considerations, nongovernmental sponsorship issues, and human-subject confidentiality protection must be considered before data are released. The NIH requires that investigators asking for funding include with their grant applications information about how they plan to share the data generated from their research. If a grant is awarded, the data-sharing plan must be enacted.

5.4 REFERENCES & ACKNOWLEDGMENT: DATA MANAGEMENT PRACTICE

1) British Medical Research Council – Recording the data (https://mrc.ukri.org/research/public-engagement/evaluating-recording/)
2) Collaborative Institutional Training Initiative – CITI Course in Data Management and Acquisition (https://www.citiprogram.org/rcrpage.asp?language=english&affiliation=100)
3) Harvard University Research Integrity Office of the Vice Provost for Research Webpage (https://vpr.harvard.edu/)
4) Harvard Office of Technology Development – Guidelines, Policies and Forms (http://otd.harvard.edu/resources/agreements/materialtransfer/#why)
5) Journal of Empirical Research on Human Research Ethics Notes, Concepts, Tools and Resources for Solving Ethical Problems in Human Research : What to Share and What to Redact: Protecting Confidentiality while Preserving Usefulness (http://www.csueastbay.edu/JERHRE/notes/DataSharing.pdf)
6) National Institute of Health - NIH Guide (https://www.nih.gov/research-training/safety-regulation-guidance)
7) National Medical Research Council – Overall Grant Framework (https://www.nmrc.gov.sg/grants)

8) National Healthcare Group Policy & Procedure Cluster Human Resources Policy – Confidentiality
9) Office of Research Integrity – Introduction to the Responsible Conduct of Research: Conducting Research, Data Management Practices (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
10) Office of Research Integrity – Introduction to the Responsible Conduct of Research: Conducting Research, Data ownership (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
11) Office of Research Integrity – Introduction to the Responsible Conduct of Research: Conducting Research, Data collection (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
12) Office of Research Integrity – Introduction to the Responsible Conduct of Research: Conducting Research, Data protection (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
13) Office of Research Integrity – Introduction to the Responsible Conduct of Research: Conducting Research, Data Sharing (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
14) Office of Research Integrity – Introduction to the Responsible Conduct of Research: Conducting Research, Future considerations (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
15) Office of Research Integrity – Policies - Statutes and Regulations (https://ori.hhs.gov/ori-introduction-responsible-conduct-research)
16) International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonized Tripartite Guideline, Guideline for Good Clinical Practice E6 (R1), Current Step 4 version dated 10 June 1996 (https://www.ich.org/page/efficacy-guidelines)
17) The Columbia University Responsible Conduct of Research : Data Acquisition and Management (http://ccnmtl.columbia.edu/projects/rcr/rcr_data/foundation/index.html)
18) The European Science Foundation – Setting Science Agendas for Europe, Member Organization Forum – Fostering Research Integrity in Europe – December 2011

19) The European Code of Conduct for Research Integrity – March 2011 (https://allea.org/code-of-conduct/)

20) The University of Kentucky Office of Research Integrity - Data Retention & Ownership Policy (https://www.research.uky.edu/research-misconduct/data-retention-and-ownership-policy)

21) The University of Oxford, University of Administration and Services (UAS), Research Data Management (http://www.admin.ox.ac.uk/researchsupport/integrity/misconduct/)