Data

1 The Nature and Recording of Data

- By data we mean recorded information, regardless of the medium of recording.

- Data may include, but are not limited to contents of notebooks, computer or instrument printouts, disks, slides, autoradiograms and questionnaire forms; among other forms of data are statistical compilations, models and their testing, and other material relevant to the research project.

- The recorded data should contain enough information to allow the researcher to reconstruct the history and details of the experiment even after the lapse of some time, to allow another researcher to replicate the research, to confirm the validity of the conclusions, to enable the researcher to respond to questions or criticisms of the findings. and to establish priority in case of a future patent application.

- In many fields of investigation, the common primary repository of experimental data is a bound notebook with consecutively numbered pages.

- Notebook entries should include a description of the methodological procedures, materials used and the nature of the observational technique, calculational and statistical treatments of the observations, results, and conclusions. Reference should be made to the location of samples or of data that cannot be recorded in the notebook.

- Each page of the notebook should be dated and initialed. When possible, printouts, tables, graphs, and photographs should be pasted into the notebook.

- Data should be recorded directly into the notebook, not through temporary slips of paper. A notebook entry should be made permanent, preferably in blue or black ink and not with red or felt-tip pens. Corrections should not be made by erasure, blackouts, or whiteouts, but by drawing a thin line through the entry to be corrected with the correcting entry above or in a nearby margin, initialed and dated, with a notation for the reason for the change, e.g., eie (error in entry).

- Refer to the Human Subjects module for recommended forms of documentation in clinical studies.

- Certain clinical studies may require the establishment of independent Data and Safety Monitoring Boards (DSMBs) that will oversee and monitor such studies to ensure the safety of participants and the validity and integrity of the data.

- NIH requires that a DSMB be established for all NIH-funded Phase III clinical trials.

- For NIH-funded Phase I and Phase II clinical trials, investigators are required to submit a general description of the data and monitoring safety plan as part of the protocol submitted to the IRB and as part of a research grant application.

- The FDA or NIH may require a DSMB for some Phase I or Phase II studies, including research on medical products intended to be marketed or studies having multiple clinical sites or involving high-risk interventions or vulnerable populations.

2 Abuse or Misuse of Data

- In general, all observed data should be included in the data analysis.

- Data should be excluded from analysis only for good reason, explained in the notebook or as part of the data analysis itself. Some acceptable reasons for exclusion might include instrumental breakdown, accidental disruption of the procedure, deterioration of an essential reagent, or irrelevance of some sections of the data, as in epidemiological studies, to the hypothesis being tested.

- Outliers, or data points that do not fit a smooth curve that goes through or near most data points, should not be excluded simply because of poor fit. Exclusions should be made only with regard to pre-defined statistical criteria included in the experimental design.

3 Confidentiality of Data

- In the case of clinical studies or surveys, primary data are associated with individuals and may convey information about the health or behavioral characteristics of individuals. The confidentiality of these data must be assured.

- A fine balance must be struck between the need to protect sensitive information that identifies individuals and the need to provide researchers with access to the information in a manner that will permit them to do their work.

- Codes should be used to identify individual research subjects, and the identification of the code name with the subject's name should be maintained in a secure place and should be available only to a small number of individuals.

- Names, addresses, birth dates, social security numbers and other identifiers should not be used in the code name.

- For further details, see the Human Subjects and HIPAA modules - 2, 6, 7 and 8.

4 Retention and Storage of Data

- Data should be stored in a safe place for as long as the research project is underway. The University has mandated a minimum retention time of five years after the final reporting or publication of the project. Since several federal agencies sponsoring or regulating research have adopted seven-year rules for data retention, the University will probably increase its mandated retention time.

- See Guidelines on Data Retention and Access, http://www.pitt.edu/~provost/retention.html.

- Beyond the five years (or longer if mandated by a sponsor), data should be retained long enough to support any patent applications and to allow the resolution of any questions about the research raised during the minimum retention period.

- Data should be suitably indexed so that they may easily be called up when necessary.

- National and international data banks exist in some fields, into which all researchers are expected to deposit their data. Some examples are X-ray crystallographic data and human genomic data. The Inter-University Consortium for Political and Social Research has prepared guidelines for preparing data for archiving. See http://www.icpsr.umich.edu/ACCESS/dpm.html

- Some journals may require that extensive data supporting a manuscript accepted for publication be deposited in an archive or at a website.

- A list of websites for social science data archives is available through the University of California at San Diego at http://odwin.ucsd.edu/idata.

5 Ownership and Access to Data

- Legal title to data acquired in University research projects rests with the University, not with the individual researcher or with the principal investigator. If any problem should be found with the data, such as an allegation that is the subject of a law suit or a research misconduct proceeding, the University has the primary responsibility to respond to the charges.

- The University may assert the right to copyright or patent products of research conducted by University researchers, in accordance with University policies and procedures dealing with intellectual property, allowing, among other features, the sharing of licensing, sale, or royalty revenues between the inventor(s) and the University.

  For Patent Policy, see:
  http://www.pitt.edu/HOME/PP/policies/11/11-02-01.html

- A researcher who has made a finding that may be patentable should file a Disclosure of Invention to the Office of Technology Management (OTM).

  For on-line filing, see the OTM website:
  http://tech-link.tt.pitt.edu

- A principal investigator who leaves the University of Pittsburgh may continue to maintain access to the data by leaving the original copy at the University and taking a copy to another institution, or by taking the original data on long-term loan with a written commitment to provide the University with access to the data at any future time within the retention period.

- If the data to be taken to another institution are governed by confidentiality restrictions, a confidentiality agreement may be executed through the Office of Research.

- Students or other investigators in a project may access those data which they have been responsible for collecting.

- The University does not normally assert its legal ownership of the literary, artistic, or scholarly work of its faculty and staff, with respect to copyright, unless the preparation of the particular work was a specific assignment as an employee of the University.

- The University does assert legal ownership to intellectual property in the form of software (other than educational software to be used only within the University) developed at least in part with the University's computer facilities. The University's ownership of software, except for software

created as a work for hire, is subject to a royalty-sharing agreement with the programmer.

- The federal government has the right of access to any data acquired under a government-sponsored research project.

- In the interest of openness, data supporting research which has been published, as well as unusual or unique materials created during the research such as DNA sequences or cell lines, should be made available to any qualified researcher who makes a reasonable request. Such professions as the American Psychology Association and the American Sociology Association have supported sharing or archiving of data.This principle has been upheld by the National Science Foundation, the National Institutes of Health and other federal agencies with respect to data generated in research supported by those agencies. An agreement should be executed for transfer of research of materials.

  For a model form for handling voluntary transfer of research materials, see: http://www.pitt.edu/~offres/proposal/mtainter.html

- Federal regulations are being developed which might require releasing to the public, in some cases, data collected in federally funded research in general and specifically when used to support federal rules and regulations, in response to a request under the Freedom of Information Act.

- There have been instances in which research data have been subpoenaed by the courts. The case law is not firmly established in this regard, and inquiries with respect to specific instances should be directed to the Office of the General Counsel.