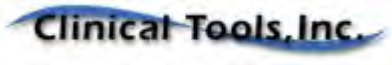
Guidelines for Responsible Data Management in Scientific Research

Developed by:



Funded by:

Office of Research Integrity
US Department of Health and Human Services

About the Course

Data management is one of the essential areas of responsible conduct of research, as outlined by the Office of Research Integrity. This educational course will educate new investigators about conducting responsible data management in scientific research. Researchers who are considering submitting a federal grant or contract for the first time can also benefit from this introductory course on data management, as can other research team members. The course includes background information about the topic, best practice guidelines, various learning features, and a resource section.

Learning Objectives

After taking the course, learners will be able to

- Understand the general rules of appropriate data management in accordance with responsible conduct of research
- Understand how to define roles and responsibilities of research staff regarding data management
- Develop and implement a communication plan for dealing with data management issues among the research team
- Utilize the information featured in the course to implement a system for conducting responsible data management

Online Version

This course was previously available on the Internet at http://www.RCREducation.com. The website is not active at this time.

Course Authors

Meghan B. Coulehan, MPH

Jonathan F. Wells, BA

Funding Information

Development of this website was funded by the Office of Research Integrity (ORI) Responsible Conduct of Research Resource Development Program.

Contact Us

Feel free to contact us with comments or questions. You can reach Project Director, Meghan Coulehan, at coulehan@clinicaltools.com.

Introduction

Data management is one of the core areas addressed by the Office of Research Integrity (ORI) in its responsible conduct of research initiative (see links in sidebar). This important, multifaceted issue affects all health researchers and deserves extra attention and diligence.

Oversight of data management represents a significant investment of time and effort by the Principal Investigator (PI) of a research project. For oversight to be thorough and correct, PIs must understand the basic concepts of data management and ensure that every member of the research project team is involved in the planning, implementation, and maintenance of data management policies and procedures.

Key Points

Data management is one of 9 core areas addressed by the Office of Research Integrity's responsible conduct of research initiative.

FYI

To learn more about the ORI or the responsible conduct in research initiative, check out the following links:

- US Department of Health and Human Services' ORI website [http://www.ori.dhhs.gov/]
- ORI's Introduction to the Responsible Conduct of Research [http://www.ori.dhhs.gov/documents/rcrintro.pdf]

Overview: Concepts of Data Management

Before starting a new scientific research project, the PI and research team must address issues related to data management, including the following:

Key Concept	How It Relates to Responsible Conduct of Research
Data Ownership	This pertains to who has the legal rights to the data and who retains the data after the project is completed, including the PI's right to transfer data between institutions.
Data Collection	This pertains to collecting project data in a consistent, systematic manner (i.e., reliability) and establishing an ongoing system for evaluating and recording changes to the project protocol (i.e., validity).
Data Storage	This concerns the amount of data that should be stored -- enough so that project results can be reconstructed.
Data Protection	This relates to protecting written and electronic data from physical damage and protecting data integrity, including damage from tampering or theft.
Data Retention	This refers to the length of time one needs to keep the project data according to the sponsor's or funder's guidelines. It also includes secure destruction of data.
Data Analysis	This pertains to how raw data are chosen, evaluated, and interpreted into meaningful and significant conclusions that other researchers and the public can understand and use.
Data Sharing	This concerns how project data and research results are disseminated to other researchers and the general public, and when data should not be shared.
Data Reporting	This pertains to the publication of conclusive findings, both positive and negative, after the project is completed.

(Steneck, 2004)

The pages that follow will provide more in-depth descriptions of each of these terms and will explain how each one relates to the responsible conduct of research.

Key Points

It is important for researchers to understand how data management issues relate to the responsible conduct of research.

FYI

You can print out the worksheet version of this page to share with your entire research team. This worksheet is included at the end of the document.

Think Ahead Quiz: What Are Data?

True or False: In scientific research, only the information and observations that are made as part of scientific inquiry are considered data.

__True
__False

Answer: False. In fact, data also include the materials, products, procedures, and other data sources that are part of the research project. Essentially, data are considered to be anything and everything that informs the way in which individuals are able to understand and to process their world. Read on to learn more.

Defining Data

Before reviewing the concepts of data management, the term **data** should be defined. The Merriam-Webster Dictionary (2005) defines data as "factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation."

According to this definition, some examples of types of medical research data would include the following:

- Patient survey responses
- White blood cell counts
- Core temperature readings
- Metabolism rates

However, data can also refer to any observations that are made -- such as a patient's symptoms or a population's health habits.

Other Forms of Data

Data are not only the information and observations made as part of scientific inquiry but also the materials, the means, and the products of that inquiry (these are sometimes called **data sources**). In other words, data can also include the following:

- Tissue samples
- Specially designed primers
- Patient questionnaires
- Interviews
- Customized online content

Key Points

Data are any information or observations that are associated with a particular project, including experimental specimens, technologies, and products related to the inquiry.

Case Vignette: Data Ownership

Dr. Smith works at The University and is the Principal Investigator on a large research project that is funded by the National Institutes of Health (NIH). However, while Dr. Smith wrote the original grant proposal, he does very little day-to-day work on the project. Instead, the Research Director, Betsy, oversees all aspects of the project, including staff supervision and all data management activities. In addition, Betsy has been lead author on several publications about the project's research findings.

Who owns the project and its data?

__ The PI, Dr. Smith
__ The Research Director, Betsy
__ The University
__ The National Institutes of Health
__ No one person or organization

Answer: The University. Despite the PI's and the Research Director's work on the project, the sponsoring institution typically maintains ownership of a project's data as long as the PI submitted the grant through that institution and is employed by them. However within the sponsoring institution, a PI is generally granted stewardship over the project data; he/she may control the course, publication, and copyright of any research, subject to institutional review. Read on to learn more about data ownership.

Data Ownership

Understanding data ownership, who can possess data, and who can publish books or articles about it are often complicated issues, related to questions of project funding, affiliations, and the sources and forms of the research itself.

For federally funded research, ownership of data involves at least 3 different entities: the sponsoring institution, the funding agency, and the PI. In many cases, the institution/organization owns the project data, but the PI and the funding agency have "rights" to access and use the data. Usually the PI has physical custody of the data on behalf of the organization. However, these rules vary by institution and depending on the funding source. Some general guidelines are presented below:

1. The Sponsoring Institution, e.g., a university or a research firm

Most often, the sponsoring institution/organization maintains ownership of a project's data as long as the PI is employed by that institution. The institution often controls all funding or the disbursement of government funding; consequently, it is also responsible for ensuring that funded research is conducted responsibly and ethically. Within the sponsoring institution, a PI is granted stewardship over the project data; the PI may control the course, publication, and copyright of any research, subject to institutional review.

2. The Funding Agency, e.g., NIH or the Centers for Disease Control and Prevention (CDC)

Many research projects are funded by federal government agencies, philanthropic organizations, or private industries. These agencies often have specific stipulations for how data will be retained and disseminated: for example, they decide whether to publish the project's results or market a resulting product, rather than the PI. The PI and institution should understand his or her funding agency's regulations regarding a research project and the data it produces. Note that requirements for federal grants may be different than government contracts (discussed further on the next page).

3. The Principal Investigator

In addition to being the steward of a project's data, a PI may retain some ownership of the data. In small businesses, it is assumed that rights and ownership of data remain with the business itself or with the funding agency, unless otherwise stipulated. In academic institutions, however, PIs are sometimes allowed to take their research and its data with them if they change research institutions. Many universities have offices and policies in place to ensure that such a transfer of data respects both the rights of the researcher and those of the institution(s).

(USDHHS, 1990)

Subjects' Rights to Ownership

It is also important to consider data ownership from the perspective of individuals who suggest research ideas and/or participate in the research. Some research subjects are expressing a desire for partial ownership or control of research in which they have participated. For instance, in 2 recent court cases, the defense contended that research institutions had improperly benefited in extending their study's implications beyond any consent that the participating subjects had given. (See sidebar for links to read more.) Since human subjects are often sources for data that may be otherwise unavailable to researchers, it is important to consider study participants' beneficence and dignity in relation to the project's progress and goals.

Key Points

Data ownership refers to the control and rights over the data as well as data management and use.

Ownership of research is a complex issue that involves the PI, the sponsoring institution, the funding agency, and any participating human subjects.

FYI

The Bayh-Dole Act of 1980 allows universities to obtain patents for inventions made with federal funding and to work directly with industry to commercialize these products. If you would like to learn more about the act's development and results thus far, follow this link to learn more about the Bayh-Dole Act.[http://www.ucop.edu/ott/bayh.html]

If you would like to learn more about the difference between government contracts and government grants, follow this link to learn about government funding through the NIH. [http://grants.nih.gov/grants/funding/contracts_vs_grants.Htm]

If you would like to learn more about how research subjects have challenged data ownership and their own role in research, read the article "Who Owns Your Genes?" from the New York Times. [http://www.nytimes.com/library/national/science/health/051500hth-aids-gene.html]

Pop up Page: Grants Versus Contracts

Much of scientific research financing from federal agencies, such as the Food and Drug Administration (FDA) or the NIH, is in the form of grants. For instance, 95% of awards that are made through NIH's Small Business Innovation Research (SBIR) program are grants, and the remaining 5% are contracts. So, what is the difference between government grants and contracts?

Government Grants

Government grants can be described as assistance funding. Grants are usually awarded to research projects that are deemed to be "good science," i.e., projects that increase our understanding of new or established theories or that further research. With a grant, the PI retains control over the scope of the research and makes decisions about how the funding will be spent.

Government Contracts

Government contracts can be described as procurement funding: that is, the government is providing money in order to acquire a product, property, or service. Like a contractual agreement between a buyer and a seller, government-contracted research is often subject to strict regulations, requirements, and expectations. For instance, the PI must coordinate project goals and decisions with the funding agency, which assigns a project officer to oversee the project and to make sure that the agency's goals are being met. Funding may be distributed in installments, contingent upon the funder's satisfaction with project progress reports. Also, the data typically belong to the funding agency, unless otherwise stipulated in the initial contract.

Think Ahead Quiz: Data Collection

Data that are collected as part of a scientific research project ultimately prove or disprove the PI's hypotheses and justify a body of research to the public at large. Which statement is true about data collection in scientific research?

__ Ensuring validity of the data is the key to successful research.
__ Ensuring reliability of the data is the key to successful research.
__ Ensuring reliability and validity are equally important.
__ Data collection is actually not a key part of scientific research, since many researchers use previously collected data.

Answer: Ensuring reliability and validity are equally important. Ensuring reliability and validity of the data are equally important during data collection. When data collection is carried out according to these 2 rules, researchers will be able to accurately assess, replicate, and disseminate their results. Read on to learn more.

Data Collection

Data collection refers not only to what information is recorded and how it is recorded, but also to how a particular research project is designed. Although data collection methodology varies by project, the aim of successful data collection should always be to uphold the integrity of the project, the institution, and the researchers involved.

Data collection may seem tedious or repetitive, but the data produced in research ultimately prove or disprove hypotheses and justify or counter a body of research. In addition, thorough data collection accomplishes the following:

- Enables those involved in the research to more accurately analyze and assess their work
- Allows independent researchers to replicate the process and evaluate results
- Impresses upon research team members the importance of data management
- Details the rationale behind a research project
- Provides justification to sponsors for expenditures and project decisions
- Yields reliable and valid results, and hypothesis testing

Key Points

Data collection provides the information necessary to develop and justify research.

A successful project collects reliable and valid data.

FYI

You can print out the worksheet version of this page to help track your data collection activities. This worksheet is included at the end of the document.

Collecting Reliable Data

Data collection guidelines and methodologies should be carefully developed before the research begins. The researchers must determine what sort of data will be collected and how this data will be analyzed. For data to be considered reliable, data collection should occur consistently and systematically throughout the course of the project.

The Importance of Planning for Data Collection

Team members who will collect data should be thoroughly trained to ensure consistency in data collection. By collecting data in a well-planned, systematic manner, team members will be able to answer any question about a project, including the following:

- The purpose behind the research
- The particular methodologies chosen
- The implementation of these methodologies
- How data that were collected and analyzed
- If unexpected results or significant errors were encountered
- The implications of the research and future directions

A clear and comprehensive account of a project and its purpose and direction make it much easier for research to be disseminated, understood, and evaluated by other members of the scientific community.

Key Points

Data collection is reliable when it is employed in a consistent and comprehensive manner throughout the course of a project.

Thorough data collection enables research team members to answer any question about a project.

FYI

For most research projects, data collection procedures are usually described briefly in grant or contract proposals. However, researchers should take the time to further define each element of data collection, including specific methodologies and plans for analysis, after receiving funding but before starting the project.

Case Vignette: Collecting Valid Data

Part of the data collection methodology for Dr. Smith's study includes distributing a 12-page self-administered questionnaire to participants; they must fill out and initial each page of the questionnaire to confirm completion.

One day on his way home from conducting an interview with a subject, the Research Assistant, Joel, needed to write directions for a friend and he reached in his bag and grabbed the first piece of paper that he could find. Joel accidentally ripped the back page off of one of the completed questionnaires to write the directions, which he then gave to his friend. He didn't realize this until a few hours later, when he was reviewing the data that he had collected that day.

Joel thought that he remembered the participant's answers on the last page of the survey, since they were mostly demographic questions.

What should Joel do?

__ Staple on a new page and fill out the subject's responses, since he remembers them.
__ Contact the subject and ask her to complete the last page of the questionnaire again.
__ Omit the participant's questionnaire from the study, his/her partial data is invalid.
__ Just pretend like he doesn't know what happened to the last page.

Answer: Omit the participant's questionnaire from the study, his/her partial data is invalid. This is Joel's best option - if he were to attempt to collect the data again from the subject, the subject would be responding in a different time and mood than when the original interview occurred. As part of responsible data management, honesty about the mishap is the best way to maintain the validity of the data and to clarify that the data were not tampered with or falsified in any way. Read on to learn more about collecting valid data.

Collecting Valid Data

Collecting valid data ensures that when research is evaluated it will be deemed good science -- meaning that the research is both precise and honest. Thorough data collection should thus include a continuous system for rigorously evaluating effective or deficient elements in the project protocol or the research team's techniques.

Record Keeping

When data are actually collected, the records should attempt to accurately represent the progress of a project and answer such questions as what, how, and why data were collected or amended. Records should be durable and accessible but safe from tampering or falsification. For smaller projects, bound notebooks provide a convenient way for all research team members to keep track of data and daily activities of a project. When keeping written records, errors should be marked and dated but never erased. This way, they can provide a quick visual account of any changes or errors that have occurred.

A downside of written records is that searching for a specific fact or trying to compare observations from several sources can be difficult. Also, maintaining handwritten records is not possible for larger projects such as clinical trials or epidemiological surveys.

More best practice tips for record keeping are provided on the next page.

Electronic Records

Electronic records allow researchers to efficiently access and compare information from different sources and across similar projects. There are numerous electronic data capture programs that allow researchers to enter, store, and audit research data. However, security of electronic records is a significant concern, although there are methods for protecting electronic records (discussed later in this course). In addition, it may be time consuming and may not be cost effective for large ongoing projects to migrate their data records to electronic files. Therefore, most projects employ a combination of written and electronic record keeping to balance the risks and benefits.

Attention to Policy and Procedure

In addition to record keeping, the validity of the data collected can also be affected by whether or not proper policies and procedures for research are followed on a project and an individual level. One should be constantly aware of all the guidelines that might apply to the project's implementation and dissemination, including special regulations that involve human and animal subjects, hazardous materials, or other controlled biological agents. Every research team member should be aware of project guidelines and standards for collecting valid data, to ensure consistency throughout the project. See the sidebar for more information and relevant links.

Key Points

Diligent record keeping is essential to ensure the validity of data.

Many research projects keep both written and electronic records in order to balance the benefits of each.

FYI

Human Subjects Research Standards

Follow this link to read the US Department of Health and Human Service's (USDHHS) Basic HHS Policy for Protection of Human Subjects. [http://www.hhs.gov/ohrp/human subjects/guidance/45cfr46.htm]

Follow this link to read the NIH's Bioethics Resources page on Human Subjects Research and Internal Review Boards (IRB). [http://www.nih.gov/sigs/bioethics/IRB.html]

Animal Research Standards

Follow this link to learn about various guidelines and issues involved in animal research from the Institute for Laboratory Animal Research. [http://dels.nas.edu/ilar_n/ilarhome/]

Follow this link to view an example of an FDA-approved protocol for testing the safety of food ingredients in animals. [http://www.cfsan.fda.gov/%7Edms/opa-pt58.html]

Pop up Page: Best Practice Tips - Record Keeping

Diligent record keeping is essential to ensuring the integrity of research data. To help maintain data validity and reliability, consider these tips when planning or completing data collection:

- **Include notes:** Your records should allow you not only to account for what occurred during the course of research but also to reconstruct and justify your findings. It is important that records include notes about what methods did or did not work, observations, and commentary on the project's progress. Keep notes according to the research team's predetermined communications plan.
- **Personal notebooks:** For smaller projects using handwritten data, each team member should have his or her own personal notebook for recording project data, observations, etc. Entries should be made in a chronological and consistent manner -- for instance, each new workday should begin on a new page. Try not to leave blank lines between entries.
- **Noting errors:** Use a consistent system for noting errors or adjustments. In written records, make entries in indelible pen so that records cannot be altered or damaged. If information needs to be changed or amended, mark through the entry with one solid line and initial and date the change. The records can thus reflect what has occurred during the course of a project.
- **Recording information:** Record anything that seems relevant to the project, its data, and the standards of the project. At a minimum, records should include the following information:
    - date and time
    - names and roles of any team members who worked with the data
    - materials, instruments, and software used
    - identification number(s) to indicate the subject and/or session
    - data from the experiment and any pertinent observations from the data's collection

    It may also be helpful to include a summary of the day's data collection activities and a task list for the next day.
- **Transferring information:** When transferring records from written to electronic format, use a double entry system to reduce rates of incorrectly entered electronic data. To implement such a system, have two different Research Assistants enter all of raw data into the software program, then cross-check the data to identify and remedy inconsistencies at the time of data entry. Use our printable worksheet to help track your data collection and entry activities. This handout is included at the end of the document.

Data Storage

Once data have been collected and recorded, the next concern is data storage. Data storage is crucial to a research project for the following reasons:

- Properly storing data is a way to safeguard your research investment.
- Data may need to be accessed in the future to explain or augment subsequent research.
- Other researchers might wish to evaluate or use the results of your research.
- Stored data can establish precedence in the event that similar research is published.
- Storing data can protect research subjects and researchers in the event of legal allegations.

Type and Amount of Data to Retain

Generally speaking, enough data should be retained so that the findings of a project can be reconstructed with ease. While this does not mean that a project needs to retain all the raw data that were collected, relevant statistics and analyses from this data should be saved, along with any notes or observations. Furthermore, if research involves the use of biological specimens, care should be taken to retain them until their quality degrades.

Electronic Data

The key issues for electronic data storage are thorough documentation to allow data to be appropriately used in the future and storage format that is easily adaptable to evolving computer hardware and software. There are some additional considerations that are unique to electronic data storage, including the following:

- Rapid access to the data
- Fast read/write rates
- Low cost
- Ability to archive the data
- Removability
- A backup system, such as storing data on CDs

(Straub, 2004)

Key Points

Storing data safeguards your research and your research investment. Storage allows future access to the data in order to re-create the findings, augment subsequent research, or establish a precedent.

Enough data should be stored so that a project and its findings can be reconstructed with ease.

Think Ahead Quiz: Data Protection

With the recent emergence of electronic databases, more scientific researchers are storing their data on their computer networks. However, data protection is an issue for both paper- and computer-based data. So what is the best way to protect data?

__ Strip identifiers from human subjects data.
__ Limit who has access to the data.
__ Use an encrypted password system and assign new passwords quarterly.
__ Destroy the written data after transferral to an electronic database.

Answer: Limit who has access to the data. This is the best way to protect data. Simple measures -- like keeping written data in a locked filing cabinet for which there is only one key -- will help minimize the chance that data could be corrupted or stolen. However, this is a complex issue and employing a multifaceted security approach is the best way to ensure that your data is protected. Read on to learn more.

Data Protection

In order to maintain the integrity of stored data, project data should be protected from physical damage as well as from tampering, loss, or theft. This is best done by limiting access to it. PIs should decide which project members are authorized to access and manage the stored data. Notebooks or questionnaires should be kept together in a safe, secure location away from public access, e.g., a locked file cabinet. Privacy and anonymity can be assured by replacing names and other information with encoded identifiers, with the encoding key kept in a different secure location. Ultimately, the best way to protect data may be to fully educate all members of the research team about data protection procedures.

How Can Data Be Protected?

Theft and hacking are particular concerns with electronic data. Many research projects involve the collection and maintenance of human subjects data and other confidential records that could become the target of hackers. In a recent example, thousands of personal information and identification records were jeopardized when hackers infiltrated systems at the University of California twice in 2005 (UTBTSC, 2005). The costs of reproducing, restoring, or replacing stolen data and the length of recovery time in the event of a theft highlight the need for protecting the computer system and the integrity of the data (Kramer et al., 2004).

Electronic data can be protected by taking the following precautions:

- **Protecting access to data**
    - Use unique user IDs and passwords that cannot be easily guessed.
    - Change passwords often to ensure that only current project members can access data.
    - Provide access to data files through a centralized process.
    - Evaluate and limit administrator access rights.
    - Ensure that outside wireless devices cannot access your system's network.
- **Protecting your system**
    - Keep updated anti-virus protection on every computer.
    - Maintain up-to-date versions of all software and media storage devices.
    - If your system is connected to the Internet, use a firewall.
    - If your system is connected to the Internet, use intrusion detection software to monitor access.
- **Protecting data integrity**
    - Record the original creation date and time for files on your systems.
    - Use encryption, electronic signatures, or watermarking to keep track of authorship and changes made to data files.
    - Regularly back up electronic data files (both on and offsite) and create both hard and soft copies.
    - Ensure that data are properly destroyed.

Third-Party Data Protection

Many research institutions have offices for information technology that work with the PI to assess the project's needs and develop a data protection protocol. For PIs without such an office, contracting with an outside information technology firm or hiring a project member to specifically focus on data protection and maintenance may be necessary. Finally, database software programs often include features that help with data protection.

Key Points

Data protection should be a part of every project's plan for data storage.

The best way to protect data, whether in written or electronic form, is by limiting access to the data.

Electronic data storage offers many benefits but requires additional consideration and safeguards.

FYI

Social engineering is a form of computer hacking in which individuals try to gain unauthorized access to computer systems and/or data in order to steal or corrupt information. Research team members need to be educated about social engineering and the importance of keeping passwords private, logging out of protected databases, and so forth.

Think Ahead Quiz: Data Retention

True or False: The USDHHS requires researchers who receive their funding to retain raw data for at least 3 years.

__ True
__ False

Answer: True. The USDHHS requires that research data be retained for a period of 3 years after the project ends. Other funding agencies have different requirements regarding data retention. Read on to learn more.

Data Retention

How Long Should Data Be Kept?

There is no set amount of time for which data should be stored. In some cases, the time period is at the discretion of the PIs; however, many sponsor institutions require that data be retained for a minimum number of years after the last expenditure report. For instance, the USDHHS requires that project data be retained for at least 3 years after the funding period ends. Other sponsors or funders may require longer or shorter periods.

Continued Storage

Once the minimum storage period has been met, the PI must decide whether to continue storing the data. Although data can be kept indefinitely, a PI must evaluate the benefits and risks of extended storage. On the one hand, one never knows when data might be needed. On the other hand, continued storage of confidential data increases the risk of possible violation. The monetary cost of retention and security are additional concerns.

Destroying Data

When the decision has been made to end data storage, data should be thoroughly and completely destroyed. Effective data destruction ensures that information cannot be extracted or reconstructed. Many document storage companies now offer onsite shredding and secure destruction of written and electronic records. For electronic data specifically, software products such as Eraser or CyberScrub are available.

Key Points

Sponsor institutions and funding agencies often have their own requirements for how long data should be retained.

Ultimately, the PI must decide when it is time to end data storage.

FYI

Learn more about data retention guidelines for the following:

NIH grants
[http://grants.nih.gov/grants/policy/nihgps/part_ii_6.htm]

A comparison of FDA, Environmental Protection Agency (EPA), and Organization for Economic Co-operation and Development (OECD) record and reporting requirements
[http://www.fda.gov/ora/compliance_ref/bimo/comparison_chart/record_report.html]

Data Analysis

Data analysis is the way raw data is chosen, evaluated, and expressed as meaningful content. For many researchers, it would be time consuming and undesirable to use **all** of the data collected over the course of a study. If it is to be translated into meaningful information, data must be managed and analyzed in an appropriate fashion.

Methods of Data Analysis

There is no single method for analyzing data. Rather, the form of analysis should come from a particular project's functions and needs. Additional considerations might include the research setting (e.g., controlled laboratory vs. field site) or the type of research (e.g., qualitative or quantitative). With few exceptions, guidelines and objectives for data analysis should be determined before a project begins.

Team Members' Responsibility

Data analysis is often delegated to a biostatistical services department (in the case of a large institutional research) or to a project's statistician. If an outside statistical service is hired to do the analysis, the PI should work with the agency to ensure that the agency understands and complies with that project's data management protocol.

While some members of the research team will be minimally involved with data analysis, they should all understand the data analysis plan and be able to interpret the results within the context of the study.

Key Points

The form of data analysis must be appropriate for the project's particular needs.

Every member of a research team should be familiar with the data analysis methods used in a project.

FYI

See next page to read more about data analysis considerations.

Pop up Page: Data Analysis Considerations

Given the important role of data analysis in a research study, it is important to avoid potential pitfalls that can invalidate or lessen the integrity of the study's data. The following are important caveats when considering the methods of analysis and the data represented:

- **Methods for analysis**
    - When planning data analyses, researchers should be aware of and work within the accepted standards for their particular area of study. Such standards include the form of data (e.g., census figures, ethnographic entries, or subject interviews) and assumptions about the populations from which the data are extracted (e.g., normally distributed or independent). If a project deviates from the accepted standards, the research team should provide justification for this deviation.
    - Significance does not imply causation or establish clinical significance or practical importance. One should be aware of the abilities as well as the limitations of a chosen method of analysis. For example, the use of subgroup analysis within a given body of data may uncover significance, both in unrecognized patterns as well as in false positives and improper correlations; further research could confirm the value of such findings.
- **Usage of data**
    - Even with an appropriate method for evaluating data, research can often run into problems over what data to include in an analysis. Common problems relating to data usage include the following:
        - whether to include or exclude outliers
        - what to do when data are missing or incomplete
        - when to appropriately alter or amend collected data
        - how to display or organize data in a meaningful way
    - Responsible data analysis attempts to accurately represent what occurred as part of the study but does not overstate the data's importance. Data analysis becomes data manipulation when finding what you want takes precedence over representing what is in the data. "Intentional falsification or fabrication of data or results" includes the following:
        - forging: inventing some or all of the reported research data or reporting experiments never performed
        - cooking: retaining only those results that fit the hypothesis
        - trimming: the unreasonable smoothing of irregularities to make the data look more accurate and precise

        (Adapted from the guidelines for integrity in research by Montana Tech at The University of Montana)

    - There are, however, instances when the amending or excluding of data is appropriate within data management:
        - after instrument problems or malfunctions
        - after loss of or change in subjects or specimens
        - after any interruptions or deviations in procedure

Case Vignette: Data Sharing

After completing the first phase of data analysis, 1 of the 3 main hypotheses of Dr. Smith and the research team was proven correct. However, the team also found some results from another facet of the project that they were not expecting. While these secondary results do not directly impact Dr. Smith's primary research questions, they may affect at least 3 other investigators' research. The results appear to be pretty definitive, but data analysis is still being conducted on other parts of the project.

The 2 Research Associates working on the project, Samantha and Enrique, are insistent that the team should immediately publish their findings in a journal, since the results may have implications on other PIs' work. Dr. Smith and Betsy, the Research Director, do not intend to publish any results for at least another year, since the research is ongoing and some questions are still unanswered.

What should the research team do?

__ They should publish the results in a journal as soon as possible.
__ They should tell the funding agency about the findings, and let the agency disseminate the information if it wants.
__ They should contact the other researchers to let them know the preliminary results.
__ They should do nothing; they aren't legally allowed to share their results until all data have been fully validated.

Answer: They should contact the other researchers to let them know the preliminary results If Dr. Smith believes that the results would have implications on other researchers' work and he does not intend to publish for quite some time, he could send his fellow researchers some information about the preliminary results as a professional courtesy and to promote collegiality. However, according to the guidelines of responsible data management, the researchers are not obligated to share their findings while the research is ongoing. Read on to learn more about data sharing and reporting.

Data Sharing and Reporting

As part of the scientific process, data are expected to be shared and reported. This serves several purposes, including the following:

- Acknowledging a study's implications
- Contributing to a field of study
- Stimulating new ideas

By sharing research results, a project may advance new techniques and theories and benefit other research. It encourages collaboration between researchers in the same field or across disciplines. Additionally, reporting of clinical research data can have a direct impact on the quality of health care provided to patients.

Data sharing usually occurs once a study has been completed. Data reporting includes discussion of the data, the data analysis, and the authorship of a project, especially in the context of a particular scientific field. Data sharing and reporting are typically accomplished by publishing results in a scientific journal or establishing a patent on a product.

Sharing Data Prior to Publication

Before publication, there is often no obligation to share any preliminary data that have been collected. In fact, sharing at this stage is sometimes discouraged because of the following reasons:

- The implications for a set of data may not be understood while a project is still in progress. By waiting until a project is ready for publication, researchers ensure that what they share has been carefully reviewed and considered.
- There is fear that less scrupulous researchers will use shared research results for their own gain. This apprehension causes some researchers to refrain from disseminating their findings (Helly et al., 2002).

However, in some cases preliminary data should be shared immediately with the public and/or other researchers since it would be of immediate benefit (e.g., if a research project found that a new drug placed subjects at grave risk or greater benefit) (Steneck, 2004). In addition, many researchers find it worthwhile to present preliminary findings in a conference setting before the study is complete to inform peers about their forthcoming research.

Sharing Data After Publication

After a project's research has been published or patented, any information related to the project should be considered open data. Other researchers may request raw data or miscellaneous information related to the project in order to verify the published data or to further their own research project. However, each project should evaluate its ability to share raw data in terms of specific needs and budget constraints.

Obligation to Report

PIs should be aware of the various guidelines and restrictions that may apply to the dissemination of their research. There are usually stipulations, specific to the funding agency or sponsor institution, describing when and how results should be shared. For instance, SBIR research may be subject to certain data reporting requirements, depending upon project phase. In addition, government-sponsored research or research related to biological agents may be subject to federal legislation such as the Patriot Act or the Freedom of Information Act.

Key Points

Data sharing is the way in which research is accurately represented to the scientific community and the general public.

Sharing information while the project is still in progress should be done cautiously, since the implications of the data may not be fully known.

Some sponsor institutions and funding agencies have their own requirements for when and how much of a research project should be shared.

FYI

The 2003 NIH policy on data sharing states the following:

"We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH endorses the sharing of final research data to serve these and other important scientific goals. The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers." Read the full text (URL below).

[http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html]

Overview: Research Team Responsibilities

Responsible data management is important in all phases of a project, from planning and data collection to data analysis and dissemination. Consequently, each research team member should know what role he or she plays in data management and his or her specific responsibilities. By clearly defining what is expected of each member and to whom each person reports, a PI can structure a project for success.

Key Points

Each member of the research team has a different role and responsibilities; these should be well defined and understood by everyone.

Think Ahead Quiz: Research Team Responsibilities

The PI is ultimately responsible for all aspects of a research project, including the oversight of data management. Which of the following tasks is usually NOT one of the PI's day-to-day responsibilities?

__ Selecting and training qualified research team members
__ Writing proposals and grant requests for a project
__ Collecting human subjects data on sensitive and confidential topics
__ Serving as a liaison to the sponsor institution
__ All of the above tasks are the PI's responsibility

Answer: Collecting human subjects data and sensitive and confidential topics. Collecting human subjects data -- even on sensitive topics -- is not usually one of the day-to-day tasks of the PI. Rather, this is usually the responsibility of a Research Assistant or sometimes a Research Associate, although there are exceptions (such as in some clinical trials, for instance). Of course, the PI is ultimately responsible for the accuracy of data collection and should be aware of the data collection protocol and progress. Read on to learn more.

Research Team Members

Although titles, roles, and responsibilities vary by organization or institution, most research teams are made up of at least 5 key members:

1. Principal Investigator

The Principal Investigator (PI) is the individual who ultimately responsible for a project and its research. The PI **enables** other team members to conduct research, and is the final authority on all scientific and medical issues related to the project. By obtaining funding and seeing that a project has the right team members, proper resources, and guidance, a PI ensures the success of the project. A project may have more than one PI, and they are Co-Principal Investigators.

2. Research Director (Project Director)

The Research Director **controls** the project. By directing the protocol for how the research and data collection are carried out, the Research Director often knows more about the day-to-day operations of the project than the PI. The Research Director works closely with the PI to both report on and redirect research.

3. Research Associate (Project Coordinator)

Under the guidance of the Research Director and the PI, the Research Associate **coordinates** the project. This individual carries out the research itself, collecting data and assessing the effectiveness of project protocol, suggesting changes to the methodology as needed.

4. Research Assistant

A Research Assistant, although normally the least experienced member of a research team, **carries out** the project work. A Research Assistant performs the day-to-day tasks of a project, including collecting and processing the data and maintaining equipment.

5. Statistician

The Statistician **analyzes** the data that are collected during the project. In some projects, the statistician may simply analyze and report on the data (under the guidance of another team member) after data collection has been completed. In other projects, a statistician is involved in the construction and analysis of research throughout the entire course of a study.

Other Team Members

Additional team members may be involved in research studies, including clinical research specialists, laboratory technicians, interns or student researchers, grant administrators, and others. Their roles should be defined by the PI at the outset of the project.

Key Points

Most research teams include at least 5 people:

1. the PI, who enables the project
2. the Research Director, who controls the project
3. the Research Associate, who coordinates the project
4. the Research Assistant, who carries out the project work
5. the Statistician, who analyzes the project data

Case Vignette: Research Team Responsibilities

After collecting data for about a year, Dr. Smith's research team revisited their original research questions. They decided to investigate an additional hypothesis related to a new issue that arose during the study. This change required adding about a dozen new questions to the self-administered questionnaire.

One day, the Research Assistant, Joel, realized that they had been administering the revised survey to subjects, but the Institutional Review Board (IRB) had not yet approved the changes.

Whose responsibility was it to make sure that data collection did not continue until the IRB approved the changes?

__ The PI, Dr. Smith
__ The Research Director, Betsy
__ The Research Associates, Samantha and Enrique
__ The Research Assistant, Joel

Answer: The Research Director, Betsy. The best answer is the Research Director, Betsy. It's true that Dr. Smith is ultimately responsible for all aspects of the project (including legal issues, as well). However, in many organizations the Research Director is responsible for day-to-day activities like ensuring that data collection does not begin or proceed unless all IRB approvals are current. Read on to learn more about specific responsibilities of research team members.

The Research Team's General Responsibilities

It is important to note that the research team members' positions may be flexible -- one person might serve in several positions or one role might involve the efforts of several individuals. Additionally, keep in mind that many organizations and/or research teams have limited funding, so team members may have to fill more than one role.

The table below provides further examples of each member's role and responsibilities, how these positions differ, and where there is overlap in team members' roles.

Team Member	Primary Responsibilities	Accountable To
Principal Investigator	• Writes grant requests and proposals for a project • Initiates a research project and aids in the design and implementation of protocols • Selects the research team members • Provides team members with the necessary technical and equipment training • Creates a structured and effective work environment • Writes and publishes research articles to disseminate project findings	• Funding agency • Sponsor institutions • Professional associations • Employer and/or contractor • Legal and academic regulations
Research Director (aka Project Director)	• Designs guidelines for project methodology, including data collection procedures • Works with PI to redefine and redirect protocol as needed • Manages team members' time and project budgetary issues • Evaluates and documents project progress and compliance with protocols • Ensures that a project complies with federal and Institutional Review Board guidelines • Assists with writing research articles to disseminate findings	• Principal Investigator
Research Associate (aka Project Coordinator)	• Follows and implements research guidelines • Coordinates and conducts experiments and data collection • Provides basic analysis for data • Monitors experiments and their compliance with the protocols • Aids in reporting project research	• Principal Investigator • Research Director • Statistician (at times)
Research Assistant	• Performs experiments and collects data • Maintains research supplies and/or equipment • Performs general background and clerical work (e.g., literature review, transcription, etc.)	• Principal Investigator • Research Director • Research Associate • Statistician (at times)
Statistician	• Ensures project design will produce reliable and valid data • Ensures research will create significant data (e.g., via sample size or analysis methods) • Monitors data collection and analysis • Analyzes and prepares data for reporting	• Principal Investigator • Research Director

Research Team Responsibilities: Data Management

Responsibilities of the PI and Research Director

Most of the specific tasks of data management fall to the PI and Research Director. For instance, these individuals are usually responsible for the following:

1. Ensuring that every person who is involved in the project knows his or her rights regarding data ownership
2. Ensuring that the protocol is meticulously planned and that staff is thoroughly trained to maintain the integrity of the data collected
3. Determining how to best store, protect, analyze, and disseminate the data
4. Developing a plan for addressing research misconduct and data mismanagement

Responsibilities of the Other Team Members

The primary data management responsibilities of the Research Associates and Research Assistants are usually in data collection: ensuring the reliable and valid collection of the data and protecting the data that they have collected. Statisticians are primarily responsible for ensuring comprehensive and accurate data analysis. All research team members are responsible for letting the PI or Research Director know if they suspect data fraud, manipulation, or other misconduct.

Key Points

The PI and Research Director are usually responsible for most of the tasks related to data management. Research Associates and Research Assistants are primarily responsible for data collection, while Statisticians are responsible for analysis.

FYI

Use our worksheet to outline each team member's responsibilities before the project begins. This worksheet is included at the end of the document.

Communication Among Research Team Members

Communication Between the PI and the Team

It is not enough for a PI to lay the groundwork for a project and then expect everything to run smoothly without any further assessment or input. After clearly defining team roles and responsibilities, a communications plan should be developed and implemented (establishing a communications plan will be discussed in the pages ahead.)

Foremost, the PI should be able to communicate well with his or her team. If possible the PI should personally educate the team members about research integrity issues, involve team members in a discussion of how data will be managed, and promote open communication amongst team members about problems or concerns. Secondly, feedback to the team is necessary. A PI's feedback keeps the team members informed about a project's developments and any changes that may directly affect individuals' roles or responsibilities. Feedback from the PI may also provide positive reinforcement. Weekly or monthly status meetings that the PI organizes and attends may help encourage feedback and open communication.

Communication Among Team Members

Similarly, team members must communicate with each other and the PI as the project progresses or when problems arise. Effective communication involves frequent and open dialogue among all team members, enabling research to proceed smoothly. A clear communications plan will ensure that everyone has an accurate picture of what is happening now and what needs to happen in the future.

Key Points

Establishing a clear and effective communications plan will ensure that all research team members are aware of the project's status, time line, changes, and any problems encountered.

Think Ahead Quiz: Communication and Leadership

A strong leader with good communications skills is able to guide both the project and the project members. Which statement is true about the role of the PI as the leader of the research team?

__ Since he or she is rarely involved in data collection or analysis, the PI defers authority to the Research Director and Statistician.
__ The PI deals with human resource issues such as benefits and paid time off.
__ The PI provides a clear, unifying vision of the project objectives, protocols, and progress to the research team.
__ The PI has minimal contact with the research team; thus, leadership is not an issue.
__ None of the above statements are true.

Answer: The PI provides a clear, unifying vision of the project objectives, protocols, and progress to the research team. The PI does serve as leader of the research team, and it is his or her role to communicate the project's vision to the research team members and to clarify each member's role and responsibilities. Read on to learn more.

The Role of Leadership in Communication

In order for a research team to function and communicate effectively, the PI must be able to lead the project and the project's members. A PI who is an effective leader conducts himself or herself as follows:

- Provides a clear vision for the project
- Defines common goals for team members
- Acts as a authority figure in the team yet is approachable
- Fosters sharing of responsibilities
- Promotes teamwork by sharing information
- Provides positive feedback and constructive criticism

Defining Common Goals

The PI must be able to provide clear project goals from the outset. However, simply providing goals does not constitute effective leadership. The PI must also unify the team by involving each team member in the vision and goals for the project. This means that the PI should make each team member aware of common goals and how that member's own role and responsibilities fit into the larger project. Defining common goals fosters motivation and accountability and promotes collaboration and communication -- individuals will know which members are responsible for what parts of a project as well as to whom each person can turn for guidance.

An Authority Figure

As the head of a project, the PI also serves as the authority figure, setting a standard for accountability and approachability that team members will rely on and replicate. Team members should feel that they can trust and approach the PI with any issues that may arise. The PI should be aware that his or her actions and decisions can affect every aspect of the project.

Managing Conflict

Given that differences are inevitable, a PI must also be able to manage conflicts among team members (discussed further on the next page).

Key Points

The PI should lead both the project and the research team by defining goals, encouraging communication and teamwork, and managing conflict.

As the head of a project, the PI also serves as the authority figure and sets the standard for accountability and approachability.

Pop up Page: Managing Conflicts Among the Research Team

Over the course of a project, it is inevitable that conflict will arise among team members. As the team's leader, the PI should be able to recognize and deal with conflict before it becomes a threat to project stability. Some potential problem areas that the PI should be aware of include the following:

- Clashing personalities between team members
- Frustration with the project or work stress
- Dissatisfaction with or refusal to follow research protocols
- Improper management of resources
- Unbalanced division of labor
- Lack of recognition or credit within a project

Regardless of the conflict's cause, its resolution must take place in an environment where team members feel they can honestly approach the PI (or another member) and express themselves. The best way to do this is by providing constructive feedback in a private setting. Constructive feedback includes the following actions:

- **Listening** to the other individual. The PI should refrain from correcting, reacting to, or otherwise interrupting the other person while he or she is speaking. The PI should engage in **active listening**, which involves demonstrating through body posture, facial expression, and attentiveness that one is aware of and interested in what the other person is trying to convey. This demonstrates respect for the other person and his or her opinions.
- **Expressing** a position in a non aggressive and nonjudgmental manner. The PI should explain and clarify the reasons behind his or her position and place these reasons in the context of the larger vision for the project or team. Expressing one's self in this way emphasizes honesty, approachability, and trust in resolving issues. Refrain from using technical jargon or expressing opinions as fact.
- **Discussing** the problem in terms of the larger picture. The PI should not critique the person but rather the idea. This means trying to understand why a particular idea is creating a conflict and uncovering any issues that could reconcile the conflict. It may be helpful to recognize and compliment the other person on some aspect of his or her idea. Doing so shows respect for the other person's opinions and demonstrates that the PI is trying to understand the logic behind it. By focusing on the conflict itself and the thought process behind it, a PI can prevent discussion from disintegrating into an argument and thus may resolve the conflict more effectively.

Case Vignette: Communication

A few weeks after Dr. Smith added the new questions to the self-administered questionnaire, it occurred to the Research Assistant, Heather, that the data collection methodology could be changed slightly. She realized that the first questionnaire that was administered to subjects (a survey on attitudes) now included information that provided answers to the questions on a subsequent questionnaire (a knowledge pre-test).

Heather realized that it would make much more sense to administer the knowledge test **before** the attitude questionnaire.

How should Heather proceed?

__ Heather should make the change with her subjects and start administering the knowledge test before the attitude questionnaire.
__ Heather should tell her fellow Research Assistants about the change so that they can all follow the same methodology.
__ Before proceeding, Heather should ask Dr. Smith for permission to make the change. Dr. Smith may have a particular reason for wanting to ask the attitude questions first.
__ Heather shouldn't do anything until she refers to the communication plan to determine Dr. Smith's system for revising the methodology.

Answer: Heather shouldn't do anything until she refers to the communication plan to determine Dr. Smith's system for revising the methodology. The research team should have a communications plan in place, and Heather should refer to this plan before she proceeds. Changes in methodology during the course of a research project are not uncommon, and it is likely that the PI has a system in place for discussing and revising the data collection procedures as needed. For instance, it may require a meeting or an e-mail or memo to affect such a change. Read on to learn more about establishing a communications system within the research team.

Establishing an Effective Communications Plan

The PI should develop and implement a communications plan at the project's outset. Whenever possible, the communications plan should be written down and distributed to all members of the research team. At any point in the project, team members should know what information is communicated, to whom, and how.

The First Steps in a Communications Plan

The first step in a communications plan is to establish the chain of command and determine who can make decisions about different aspects of the project. Basic ground rules also should be outlined, such as whether or not the team should keep written or electronic records of important communications. A good communications system will serve as a check-and-balance system and maintain the integrity of the research project.

Best practice tipe for communication are discussed further on the next page.

The Next Steps

The communication plan should also address data collection issues. A system for monitoring and checking data collection should be defined well before data collection begins. Such a system should document each step in the data collection process and whose responsibility it is. The following questions should be addressed:

- How much data have been collected and by whom?
- Have the data been entered or transferred into an electronic format?
- Have the transferred data been double-checked against the original (by a different team member) to ensure accuracy?
- For human subjects data, have identifiers been stripped from each and every record?

Other Data Management Issues to Consider

The communications plan should consider other data management activities as well. For example, while the PI and Research Director don't need to be informed every time a Research Assistant collects new data, the communications plan should outline how the Research Assistant updates the team. In this instance, the Research Assistant could provide a weekly e-mail to the team with a summary of data collection activities, or he or she could log daily activity in a notebook.

Another example of a communications issue to be considered is how a team member might convey the results of a monthly virus scan on the entire network. The plan might require the Research Associate to keep a logbook, with dated entries for each scan that is run without incident. The communications plan should be also deal with a scan that finds a potentially harmful computer virus.

Key Points

The PI should establish and implement a communications plan at the start of the project; all research team members should receive a written copy of the plan.

Data management activities and progress should be included in the communications plan.

FYI

Communication can be conceptualized as more than just written and verbal. The PI should also consider the role of the following:

- Internal (within the research team) and external (other project stakeholders) communications
- Formal (reports, grant proposals) and informal (memos, e-mails) communications
- Vertical (within the research team) and horizontal (between peers) communications

(Project Management Institute, 2000)

Pop up Page: Best Practice Tips: Communication

Establishing a communications plan will help the project run more smoothly. When starting a new project, consider these best practice tips on research team communication:

- Create a flowchart that lists all members of the research team, their responsibilities, who they are accountable to, who they supervise, etc. Include this in the communication plan or post it in a common area.
- Develop a plan for reporting project progress, proposed changes, and problems. An e-mail or memo may suffice for some issues, while other issues may require a team meeting.
- Hold team meetings on a regular basis as well as one-on-one meetings with individual team members. These conversations provide an opportunity for members to provide feedback or bring up problems that they might not feel comfortable discussing in front of the whole team.
- Create a team calendar that contains important dates for your project, such as team meetings or deadlines for progress reports. In addition, choose a way to notify team members, perhaps via e-mail or memorandum, when important dates are approaching.
- Clearly outline rights to data ownership, intellectual property, and publication when a project is collaborative or involves the efforts of several PIs and/or Research Directors. Specify how and when research data can be published so as to avoid confusion later on.
- Even if not required, consider establishing a structured system for communicating with the sponsor institution and the funding agency. This may entail making periodic phone calls or sending monthly progress reports to keep them informed about the status of the project.

Conclusion

Data management is a critical component of most scientific research studies. The PI should consider the following issues when establishing a data management system for a new research project. Addressing each of these issues at a project's inception will allow the PI to run an organized research project.

Issue to Be Addressed	Action to Take
Data Management Needs and Preferences	After outlining the project needs regarding data collection, storage, protection, retention, etc., the PI should assign tasks related to each of these needs to the appropriate team member.
Research Team Members' Skills and Experience	The PI should be familiar with each team member's skills so that appropriate tasks can be assigned and/or training can be arranged when needed.
Research Team Members' Roles and Responsibilities	The PI should clearly define each team member's responsibilities for each aspect of the project so that the data's integrity is maintained at all times.
Potential Problems and Solutions	At the start of the project, the PI should review other data management issues -- such as those related to data ownership and sharing -- to determine if they pose a concern.
Project Time Line	After establishing an action plan for completing the project, the PI should write a detailed time line, to keep the entire team informed of important dates and deadlines.

Key Points

The PI should consider the project's data management needs, the research team members' skills and experience, the project's time line, and potential problems and solutions when starting a new project.

FYI

Use our worksheet to outline each team member's skills and responsibilities at the start of a new project. The worksheet is included at the end of the document.

Review of Key Points

Basics Concepts in Data Management

Data management includes several key concepts. It is important to understand what these terms mean as well as how they relate to the responsible conduct of research.

- Data are any information or observations that are associated with a particular project, including experimental specimens, technologies, and products related to the inquiry.
- Data ownership refers to the control and rights over the data as well as data management and use. Data ownership is a complex issue involving the PI, the sponsoring institution, the funding agency, and any participating human subjects.
- Data collection provides the information necessary to develop and to justify research. A successful project collects reliable and valid data. Data collection is reliable when it is employed in a consistent and comprehensive manner throughout the course of a project.
- Diligent record keeping -- whether written or electronic -- is essential to ensure the validity of data.
- Storing data safeguards a research investment. Storage allows future access to the data in order to re-create the findings, augment subsequent research, or establish a precedent. Enough data should be stored so that a project and its findings can be reconstructed with ease.
- The best way to protect data is to limit access to it, whether the data are in written or electronic form. Electronic data storage requires additional safeguards.
- Sponsor institutions and funding agencies often have their own requirements for data retention; ultimately, the PI must decide when it is time to end data storage.
- Data analysis of a project must be appropriate for the project's particular needs.
- Data sharing while a project is still in progress is often discouraged, since the implications of the data may not be fully known. Some sponsor institutions and funding agencies have their own requirements for when and how much of a research project should be shared.

Research Team Responsibilities

Each member of the research team has a different role and responsibilities; these should be well defined and understood by everyone.

- Most research teams include at least 5 people: the PI, who enables the project; the Research Director, who controls the project; the Research Associate, who coordinates the project; the Research Assistant, who carries out the project work; and the Statistician, who analyzes the project data.
- The PI and Research Director are usually responsible for most of the tasks related to data management. Research Associates and Research Assistants are primarily responsible for data collection, while Statisticians are responsible for analysis.

Establishing a Communications Plan

Establishing a clear and effective communications plan will ensure that all research team members are aware of the project's status, time line, changes, and any problems encountered.

- The PI should lead both the project and the research team by defining goals, encouraging communication and teamwork, and managing conflict. As the head of a project, the PI also serves as the authority figure and sets the standard for accountability and approachability.
- The PI should establish and implement a communications plan at the start of the project; all research team members should receive a written copy of the plan, which should also address data management activities.

Final Step

Thank you for viewing our data management course! The following references and useful resources are included below:

- Course References

Resources

- Data Management – General
- Data Ownership and Retention
- Data Collection and Record Keeping
- Data Storage and Protection
- Data Sharing and Publication
- Human Subjects Research
- Animal Research
- Research Team Leadership and Communication

References

Alemi F, Maddox PJ, Prudius V, Doyon V. Evaluating Medicaid HMOs when encounter data are missing: case of developmentally delayed children. Health Care Management Science. 2003;6(1):37.

American Society of Mechanical Engineers. Welcome to ASME professional practice curriculum. Available at: http://www.professionalpractice.asme.org/index.htm. Accessed September 6, 2005.

American Statistical Association. Ethical guidelines for statistical practice. August 7, 1999. Available at: http://www.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics. Accessed August 1, 2005.

Bierig JR. Informed consent in the practice of pathology. Archives of Pathology and Laboratory Medicine [serial online]. 2001;125(11):1425–1429. Available at: http://arpa.allenpress.com/arpaonline/?request=get-document &doi=10.1043%2F0003-9985(2001)125%3C1425:ICITPO%3E2.0.CO%3B2. Accessed September 20, 2005.

Council on Governmental Relations. The Bayh-Dole Act: a guide to the law and implementing regulations. October 1999. Available at: http://www.cogr.edu/docs/Bayh_Dole.pdf. Accessed August 5, 2005.

Council for Responsible Genetics. Genetics and the law: Greenberg v. Miami Children's Hospital. Available at: http://www.genelaw.info/pages/casedetail.asp?record=46. Accessed on August 5, 2005.

Council for Responsible Genetics. Genetics and the law: Moore v. Regents of the University of California. Available at: http://www.genelaw.info/pages/casedetail.asp?record=8. Accessed on August 5, 2005.

Data. Merriam-Webster Online Dictionary. 2005. Available at: http://www.m-w.com. Accessed August 1, 2005.

Duke University. Institutional Review Board. Clinical investigations procedures and guidelines for the protection of human research subjects. Available at: http://irb.mc.duke.edu/guide.htm. Accessed August 8, 2005.

Foss B, Henderson I, Johnson P, Murray D, Stone M. Managing the quality and completeness of customer data. Journal of Database Management. 2002;10(2):138-158.

Georgetown University. An overview of biostatistics. Available at: http://www.georgetown.edu/research/arc/biostat2.html. Accessed August 8, 2005.

Geringer JM, Frayne CA, Milliman JF. In search of best practices in international human resource management: research design and methodology. Human Resource Management. 2002;41(1):5-30.

Gorner P. Parents suing over patenting of genetic test. Chicago Tribune [serial online]. November 19, 2000. Available at: http://home.iprimus.com.au/dna_info/dna/JA_DNA_ChiTrib_20001119a.html. Accessed September 20, 2005.

Gottlieb S. Opening Pandora's box: using modern information tools to improve drug safety. Health Affairs. 2005;24 (4):938-948.

Ha WT, Morris RK. SPC for nonstatisticians. Quality. 2003;42(6):42.

Helly JJ, Elvins TT, Sutton D, et al. Controlled publication of digital scientific data. Communications of the ACM. 2002;45(5):97-101.

Institute for Laboratory Animal Research. Institute for Laboratory Animal Research home page. Available at: http://dels.nas.edu/ilar_n/ilarhome/. Accessed on August 8, 2005.

Kolata G. Who owns your genes? New York Times [serial online]. May 15, 2000. Available at: http://www.nytimes.com/library/national/science/health/051500hth-aids-gene.html. Accessed September 20, 2005.

Koslowsky S. The case of the missing data. Journal of Database Management. 2002;9(4):312-318.

Kramer WTC, Shoshani A, Agarwal DA, et al. Deep scientific computing requires deep data. IBM Journal of Research and Development. 2004;48(2):209-232.

Marsh R. Drowning in dirty data? It's time to sink or swim: a four-stage methodology for total data quality management. Journal of Database Marketing and Customer Strategy Management. 2005;12(2):105-112.

Michigan State University. Transferring research. Available at: http://www.msu.edu/unit/vprgs/transferringres.htm. Accessed August 8, 2005.

References (continued 2)

Montana Tech at The University of Montana. A policy to assure the integrity of research and scholarly activity. February 14, 2000. Available at: http://www.mtech.edu/research/proposalprep/MT%20Research%20Integrity%20Policy%20at%203-2-00.pdf. Accessed August 15, 2005.

Mullner R, Chung K. The American Hospital Association's annual survey of hospitals: a critical appraisal. Journal of Consumer Marketing. 2002;19(7):614-618.

National Institutes of Health. Clinical research training course. Available at: http://www.cc.nih.gov/researchers/training/crt.shtml. Accessed August 8, 2005.

National Institutes of Health. Contracts vs. grants: what's the difference? Available at: http://grants.nih.gov/grants/funding/contracts_vs_grants.htm. Accessed August 8, 2005.

National Institutes of Health. Human subjects research and IRBs -- bioethics resources on the Web. February 1, 2005. Available at: http://www.nih.gov/sigs/bioethics/IRB.html. Accessed September 10, 2005.

National Institutes of Health. NIH data sharing policy and implementation guidance. March 5, 2003. Available at: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. Accessed August 8, 2005.

National Science Foundation. Directorate for Engineering: DMII reporting information for SBIR. Available at: http://www.nsf.gov/eng/sbir/reporting2.jsp. Accessed August 8, 2005.

Pennsylvania State University. Building blocks for teams. June 24, 2005. Available at: http://tlt.its.psu.edu/suggestions/teams/student/index.html. Accessed on September 6, 2005.

Pons AP, Aljifri H. Data protection using watermarking in e-business. Journal of Database Management. 2003;14(4): 1-13.

Project Management Institute. A Guide to the Project Management Body of Knowledge. Newtown Square, Penn: Project Management Institute; 2000.

Rhodes LJ. Institutional environments and responsible conduct of research (RCR) [University of Nevada - Las Vegas Web site]. Available at: http://www.unlv.edu/Research/icoi/Presentation%205%20-Institutional%20Environments%20 &%20RCR%20-%20Las%20Vegas%20-%2012-04%20-%20L%20Rhoades.pdf. Accessed August 1, 2005.

Sigma Xi The Scientific Research Society. 2000 forum proceedings: oversight of research staff by principal investigator. Available at: http://www.sigmaxi.org/meetings/archive/forum.2000.oversight.shtml. Accessed August 1, 2005.

Society for Clinical Data Management. Good clinical data management practices, version 3. 2003. Available at: http://www.scdm.org/gcdmp/gcdmp_v3_e04350d030g12111.pdf. Accessed on August 8, 2005.

Steneck NH. Introduction to the responsible conduct of research [Office of Research Integrity Web page]. 2004. Available at: http://ori.dhhs.gov/documents/rcrintro.pdf. Accessed August 1, 2005.

Straub J. The digital tsunami: a perspective on data storage. Information Management Journal. 2004;38(1):42-50.

Tonkens R. Clinical research organizations offer wide range of management opportunities for physician executives. Physician Executive. 2005;31(1):38-40.

University of Alaska - Fairbanks, Office of Research Integrity. Research policies. Available at: http://www.uaf.edu/ori/res_policies.htm. Accessed August 1, 2005.

University of California - Berkeley, Office of Human Resources. Guide to managing human resources: a resource for managers and supervisors at Berkeley. Available at: http://hrweb.berkeley.edu/guide/contents.htm. Accessed on September 6, 2005.

University of California - San Francisco, Office of Research. New investigators: a quick guide to starting your research at UCSF. Available at: http://www.research.ucsf.edu/QG/orQgDm.asp#Points. Accessed August 1, 2005.

University of Florida, College of Medicine. GMS 6931: responsible conduct of biomedical research. Available at: http://idp.med.ufl.edu/rcr/. Accessed on August 8, 2005.

References (continued 3)

University of Louisville. Research integrity program: guidance for development of a management plan. Available at: http://www.ori.louisville.edu/SFI/Management_Plans/Management.htm. Accessed on August 1, 2005.

University of New Hampshire. Policy on ownership and management of research data. Available at: http://www.unh.edu/orps/downloads/DataOwnershipandManagement.pdf. Accessed August 8, 2005.

University of North Carolina - Chapel Hill. Office of Human Research Ethics page. Available at: http://research.unc.edu/ohre/. Accessed August 8, 2005.

University of Pittsburgh. Guidelines on data retention and access. Available at: http://www.pitt.edu/~provost/retention.html. Accessed August 8, 2005.

University of Pittsburgh. Office of Research Integrity Guidelines for Ethical Practices in Research page. Available at: http://www.pitt.edu/~provost/ethresearch.html. Accessed on August 8, 2005.

University of Texas - Brownsville, Texas Southmost College, Corporate Compliance Office. Compliance corner. April 13, 2005. Available at: http://www.nih.gov/sigs/bioethics/IRB.html. Accessed September 10, 2005.

US Dept of Health and Human Services. Data management in biomedical research: report of a workshop. Presented at: Workshop on Data Management in Biomedical Research; April 25, 1990; Chevy Chase, Md.

US Dept of Health and Human Services. Guidelines for the conduct of research within the Public Health Service [East Tennessee State University Web site]. 1992. Available at: http://www.etsu.edu/research_ethics/ori.guidelines.pdf. Accessed August 1, 2005.

US Dept of Health and Human Services. OHRP code of federal regulations: protection of human subjects. Available at: http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm. Accessed September 10, 2005.

US Food and Drug Administration. Comparison chart of FDA and EPA: records and reports. Available at: http://www.fda.gov/ora/compliance_ref/bimo/comparison_chart/record_report.html. Accessed September 10, 2005.

US Office of Personnel Management. GS-2200 –– information technology group. Available at: http://http://www.opm.gov/fedclass/text/GS-2200.htm. Accessed August 8, 2005.

Wake Forest University. Policies and procedures: research related policies –– ethical standards in research. Available at: http://www1.wfubmc.edu/OR/Policies+and+Procedures/Manual/Section+1+-+Ethical+Standards.htm. Accessed on August 8, 2005.

Winkler A, McCarthy P. Maximising the value of missing data. Journal of Targeting, Measurement and Analysis for Marketing. 2005;13(2):168-178.

Online Resources

General Data Management

Northwestern University: Policies & Guidelines for Investigators in Scientific Research. http://www.northwestern.edu/research/policies/investigatorsIntegrity.html

- This website includes an explanation of research misconduct and research integrity as well as guidelines specifically for Northwestern staff that can be adapted to other research settings.

Office of Research Integrity: Introduction to the Responsible Conduct of Research. http://ori.dhhs.gov/documents/rcrintro.pdf

- This ORI publication provides a brief overview of the 9 core concepts related to responsible conduct of research.

University of California - San Francisco, Department of Neurological Surgery: Guidelines on Research Data and Reports. http://neurosurgery.medschool.ucsf.edu/academics/guidelines.html

- This online document describes "good research practices" for PIs, including guidelines for data management, record keeping, authorship, and data reporting.

Yale University School of Medicine, Office of Grant & Contract Administration and Scientific Affairs: Guidelines for the Responsible Conduct of Research at Yale University School of Medicine. http://grants.med.yale.edu/policies/guidelin.html

- This resource outlines the research policies and guidelines at Yale University on topics such as research team responsibilities, data management, and data ownership/authorship.

Data Ownership and Retention

Office of Management and Budget. Circular No. A-110: Uniform Administration for Grants. Section 53: Retention and access requirements for records. http://www.whitehouse.gov/OMB/circulars/a110/a110.html#53

- This circular from Executive Office of the President describes the legal retention and access requirements for records from federally funded research.

University of Arizona: Handbook for Principal Investigators -- Technical Responsibilities. http://www.sps.arizona.edu/handbook/technicalresponsibilities.htm

- This section from the University of Arizona's Handbook for Principal Investigators describes requisite technical responsibilities for the position, including concerns for data ownership, retention, and changes to research protocol.

University of Chicago: University Research Administration -- Regulations, Policies, and Procedures. Intellectual Property, Data Rights, and Data Retention. http://researchadmin.uchicago.edu/regulations/intellectual_prop.shtml

- This website from the University of Chicago discusses various policies on intellectual property and data rights as well as a section containing helpful links on these topics.

United States Copyright Office. http://www.loc.gov/copyright

- This is the home page of the U.S. Copyright Office; it contains information on registering and searching for copyrights.

United States Patent and Trademark Office. http://www.uspto.gov

- This is the homepage of the U.S. Patent and Trademark Office and it contains information on filing

and searching for currently registered patents.

Data Collection and Record Keeping

University of Florida, Office of Technology Licensing: Good Record Keeping -- Procedures for Academic Laboratory Settings. http://rgp.ufl.edu/otl/goodrecords.html

- This website from the University of Florida describes both the need for and the implementation of successful record keeping in academic laboratory settings.

University of California - San Francisco, Office of Research: New Investigators Quick Guide: Guidelines for Laboratory Notebooks. http://www.research.ucsf.edu/QG/orQgNb.asp

- This section from UCSF's New Investigators Quick Guide describes how to properly keep and maintain laboratory notebooks.

Commonwealth of Australia, National Archives of Australia: Digital Recordkeeping Guidelines -- Guidelines for Creating, Managing, and Preserving Digital Records. http://www.naa.gov.au/recordkeeping/er/guidelines.html

- This web page from the National Archives of Australia provides a comprehensive set of guidelines for digital record keeping, including issues related to creation, storage, protection, and destruction.

University of Michigan, University Archivists Group: Electronic Recordkeeping Guidelines. http://www-personal.umich.edu/~deromedi/CIC/guide.htm

- This web page contains links to guidelines and other online resources for electronic record keeping used by the University Archivists Group at the University of Michigan.

Data Storage and Protection

Economic and Social Data Service: Identifiers and Anonymisation: Dealing With Confidentiality. http://www.esds.ac.uk/aandp/create/identguideline.asp

- This web page discusses the proper way to remove or to restructure research identifiers in order to maintain confidentiality.

University of Bath: General Data Protection Guidelines for Staff and Students. http://internal.bath.ac.uk/data-protection/guidelines.htm

- This website from the University of Bath discusses general data protection guidelines, highlighting 8 principles for achieving successful data protection compliance.

University of Minnesota, Institutional Review Board: Electronic Data Storage and Security. http://www.research.umn.edu/irb/guidance/data/index.cfm

- This section of the University of Minnesota's Guidance for Research provides recommendations for keeping research data secure, including tips for passwords and links to data security products.

Data Sharing and Publication

Harvard University: Data Sharing and Replication. http://gking.harvard.edu/replication.shtml

- This website contains a wide range of links on data sharing, including discussions and relevant policies for various journals and funding agencies.

Mount Sinai School of Medicine: Handbook for Research -- Section III: Guidelines for Reporting Research Results. http://www.mssm.edu/forfaculty/handbook_rs/articles.shtml

- This website from Mount Sinai School of Medicine provides guidelines for submitting articles to

scientific journals and discusses what constitutes appropriate content, citation, and authorship.

National Institutes of Health: NIH Data Sharing Policy. http://grants.nih.gov/grants/policy/data_sharing

• This website contains information, FAQs, workbooks, and testimonials on the subject of data sharing in relation to the NIH's data sharing policy.

Online Ethics Center: Research Ethics Module. Responsible Authorship.
http://onlineethics.org/reseth/mod/auth.html

• This module from the Online Ethics Center discusses responsible authorship, providing scenarios and suggested readings on the subject.

Human Subjects Research

Centers for Disease Control and Prevention, Office of the Chief Science Officer: Human Subjects Documents.
http://www.cdc.gov/od/ads/hsrdocs.htm

• This website from the CDC contains a variety of documents on the subject of human subjects research, including guides for writing consent documents, responding to IRB reports, and protecting research subjects.

National Institutes of Health: Human Subjects Research and IRBs. http://www.nih.gov/sigs/bioethics/IRB.html

• This resource page from the NIH contains links to information on human subjects research, among them links to policies and regulations, IRB resources, and guidance for investigators.

National Institutes of Health, Office for Protection from Research Risks: 1993 Institutional Review Board Guide -- Protecting Human Research Subjects. http://www.genome.gov/10001752

• This 1993 guidebook from the NIH explains and discusses the issues involved in approving and reviewing human genetic research by Institutional Review Boards.

United States Department of Energy, Office of Biological and Environmental Research: Protecting Human Subjects. http://www.er.doe.gov/production/ober/humsubj/index.html

• This website from the Department of Energy contains resources for human subjects research, including a project database, consent form information, and details on receiving accreditation.

US Department of Health and Human Services, Office of Human Research Protections.
http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm

• This website contains the Code of Federal Regulations, as set up by the USDHHS-OHRP for the protection of human subjects.

Animal Research

Institute for Laboratory Animal Research. http://dels.nas.edu/ilar_n/ilarhome

• This website provides science-based guidelines for animal research as well as information on various animal models and strains.

National Institutes of Health, Office of Laboratory Animal Welfare. http://grants.nih.gov/grants/olaw/olaw.htm

• The Office of Laboratory Animal Welfare website provides links to current news flashes, policies and laws, guidance, and other resources within the realm of animal research.

Rutgers University School of Law: Animal Rights Law Project -- Federal Animal Welfare Act and Regulations.

http://www.animal-law.org/welfact

- This website contains information on the United States code and its regulations that govern the treatment and handling of animals in research and nonresearch settings.

United States Department of Agriculture, National Agricultural Library: Animal Welfare Information Center. http://www.nal.usda.gov/awic/index.html

- The Animal Welfare Information Center website provides a variety of information on subjects from government and legal resources on lab animals, zoos, circuses, and wildlife.

Research Team Leadership and Communication
Dartmouth University, Office of Sponsored Projects: Role of the Principal Investigator. http://www.dartmouth.edu/~osp/resources/manual/post-award/pirole.html

- This webpage describes the role of the Principal Investigator in sponsored research.

Sigma Xi The Scientific Research Society: 2000 Forum Proceedings -- Oversight of Research Staff by Principal Investigator. http://www.sigmaxi.org/meetings/archive/forum.2000.oversight.shtml

- This panel discussion attempts to describe how Principal Investigators should manage their research staff, citing cases of research misconduct and case scenarios.

University of California, Office of Human Resources: Guide to Managing Human Resources -- A Resource for Managers and Supervisors at Berkeley. http://hrweb.berkeley.edu/guide/contents.htm

- Berkeley's Guide to Managing Human Resources contains a wide range of information on subjects such as recruiting staff, managing staff successfully, and promoting successful work relations.

Review of Key Concepts in Data Management

	How it Relates to Responsible Conduct of Research
Data Ownership	Concerns who has the legal rights to the data and who retains the data after the project is completed, including the PI's right to transfer their data between institutions
Data Collection	Concerns collecting data in a consistent, systematic manner throughout the project (reliability) and establishing an ongoing system for evaluating and recording changes to the project protocol (validity)
Data Storage	Concerns the amount of data that should be stored - enough so that project results can be reconstructed
Data Protection	Concerns protecting both written and electronic data from physical damage as well as damage to data integrity, including tampering or theft
Data Retention	Concerns how long project data needs to be retained according to various sponsors' and funders' guidelines, and the importance of secure destruction of data
Data Analysis	Concerns how raw data is chosen, evaluated, and interpreted into meaningful and significant conclusions that other researchers and the public can understand and use
Data Sharing	Concerns how project data is disseminated to other researchers and the general public to share important or useful research results; also, when data should not be shared
Data Reporting	Concerns publication of conclusive findings after the project is completed

For more information about Responsible Conduct of Research, visit the Office of Research Integrity's website at http://ori.dhhs.gov

Project _____ **Page#** _____

DOUBLE-ENTRY

Data Collection			Data Transfer #1		Data Transfer #2		PI/RD Review	
		Staff Initials		Staff Initials		Staff Initials		Staff Initials

Project _____

Team Member	Skills & Strengths		Other Responsibilities
Name: Supervised by: Supervisor to:			
Name: Supervised by: Supervisor to:			