CHAPTER

8

Data Management and Research Integrity

Shannon Wapole, David A. Stone
Northern Illinois University, DeKalb, IL, USA

1. INTRODUCTION

Data management has always been central to the entire scientific research enterprise. Scientific progress depends on the ability to reproduce, extend, and challenge scientific findings. These activities all require that research, whether experiments or field studies, be properly documented in their procedures, processes, and findings. The scientific community further requires that these documents and other artifacts of scientific discovery be properly stored and managed to maintain their propriety, and that they be functionally accessible. In these ways, data management is understood to be central to the responsible conduct of research (RCR).

But over the past decade, the need for data management has taken on a new character and with it a new urgency. The capacity to produce and reduce scientific data to digital form has ushered in a completely new scientific era from the standpoint of data management. The dawning of this era was recognized in the United States by the National Science Foundation (NSF) at the turn of the new century and heralded by the release of two reports in 2003 that recognized the urgency of addressing the growing need for what was termed "cyberinfrastructure" for the collection, storage, archiving, and sharing of the vast quantities of digital data that science was producing and expected to produce in the future [1]. A decade later, the White House issued an executive order calling on all of the major federal research funding agencies to put forward more robust plans that allow for the retention and sharing of research data and the public availability of federally funded research findings [2]. It is therefore important for researchers, research office personnel, and library and information technology staff to understand the general rules of data management, the roles of researchers in regard to data management, the need to develop and implement data management plans, and how to ensure compliance with data management regulations.

In 2003, the National Institutes of Health (NIH) also responded to the need to address calls for data management, data sharing, and increased access to research results by issuing a data

Research Regulatory Compliance
http://dx.doi.org/10.1016/B978-0-12-420058-6.00008-3

sharing policy and requiring applicants for large grants to submit data management plans with their applications [3]. Since then, the regulatory environment for data management is undergoing constant change. As big data and collaboration become scientific watchwords, regulations directly addressing data management are evolving, and are expected to become more comprehensive with time. Most federal agencies that fund the creation and use of data already have data management regulations, but as the nature of technology and the ways we use data change, these requirements can be expected to change to keep pace.

This chapter will provide researchers and research office personnel with a basis for understanding the current regulatory environment and its likely trajectory following the latest executive orders on data management of open access. In addition, because data management is now a research compliance matter, the chapter will also provide information on common compliance challenges and the role data management practices play in issues related to research integrity and scientific misconduct. As such, this chapter does not address other matters, such as the development of national and international data management tools, repositories, and other infrastructure supports, issues related to resources for implementation of data management plans, the need for training in the area of data management, or other issues that are important to addressing data management.

2. HISTORICAL PERSPECTIVES

2.1 Current Changes in Federal Agencies

2.1.1 National Institutes of Health

The federal focus on issues related to data management arose as a confluence of matters relating to the safeguarding of the collection of unique and costly data; the need for data sharing to promote replication, extension, and innovation; and the desire to ensure that results of scientific research be made as widely available as possible to both the research community and the public at large. The NIH was the first institution to respond to these calls with regulations regarding data sharing [3]. According to NIH regulations, applications to NIH for funds exceeding $500,000 were required to include a management plan to promote the sharing of research data. Additionally, recipients of NIH funding were required to submit their papers to PubMed Central, a data repository where all publications are free to the public. A statement of this regulation promulgated in 2008 states: The Director of the NIH shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication: Provided that the NIH shall implement the public access policy in a manner consistent with copyright law [4].

2.1.2 Multiagency Task Force

In January 2009, The Working Group on Digital Data issued a report, Harnessing the Power of Digital Data for Science and Society, calling for the promotion of data management plans [5]. This task force was represented by 22 federal agencies. The report contains a vision, calling for "a digital scientific data universe in which data creation, collection, documentation, analysis,

preservation, and dissemination can be appropriately, reliably, and readily managed" [5]. The report also called for agencies to stress the importance of making data readily available to the public. Specific suggestions for data management plans were also outlined. In the wake of this report, federal agencies have already made changes regarding their data management policies.

2.1.3 National Science Foundation

In 2010, the NSF changed its requirements for all grant applications submitted after January 17, 2011 [6]. Under the change in regulations, all proposals must include a data management plan. This plan must include information on how the applicant will conform to NSF's data dissemination and sharing policy, which states: "Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections, and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing" [6].

2.1.4 Office of Science and Technology Policy

In February 2013, the White House issued a new directive under the Office of Science and Technology Policy [7]. A memo accompanied the update regarding increasing access to the results of federal funded scientific research. Funding agencies with an annual research and development budget greater than $100 million dollars must develop a public access plan for distributing the results of their research [8].

Part of this requirement involves data management plans. "Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans and, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified" [8].

It is expected that other federal agencies will develop policies requiring grant applications to include information regarding the collection, storage, protection, retention, analysis, sharing, and reporting of scientific data. These regulations will evolve with time and it is important that researchers become familiar with creating and implementing data management plans.

2.2 Data Management Defined

Historically, data management has referred to data ownership and sharing. However, data management has a much more expansive definition and contains many facets. Following is a brief outline of the different types of data management.

2.2.1 Data Collection

Data collection refers to what information will be recorded, how it will be recorded, and the design of any particular research project. Data collection should be planned beforehand and used consistently throughout the project. This is the best way to ensure the integrity of the data [9].

2.2.2 Data Analysis

Determining what raw data should be analyzed and how the data will be analyzed is data analysis. This is an important aspect of data management for researchers to consider because analyzing the wrong data could lead to false results [10].

2.2.3 Data Storage

It is necessary for the research to determine what, how, and where the data should be stored. Certain projects can yield a large amount of data, not all of which is valuable. Also, some data may need to be stored in specific conditions. When dealing with electronic data storage backup files should be created [9].

2.2.4 Data Retention

Many funding organizations have time requirements for retaining data. Institutions may also have restrictions on timelines for retaining data. Researchers should be aware of the timeframe prior to beginning research [11]. Data retention also includes proper destruction of data, such as making sure that all confidentiality laws are being followed [9].

2.2.5 Data Ownership

Data ownership concerns who owns the data. This can extend to laboratory notebooks, notations, and hard data. Data ownership questions extend beyond the researchers tenure with the institution where the research was performed [12].

2.2.6 Data Protection

It is necessary that the written and electronic data are protected from physical damage, tampering, or theft [11]. This may include encrypting digital data.

2.2.7 Data Sharing

Data sharing is one of the most commonly discussed issues regarding data management. This pertains to how the project data and research results are made available to the general public [10].

2.2.8 Data Reporting

After the research project is done, the results of the data should be published. This is not limited to positive results and is likely the final step in data management [9].

3. RELEVANT REGULATORY/OVERSIGHT AGENCIES, REGULATIONS, AND GUIDANCE DOCUMENTS

Following the changes in NSF in regulations, researchers at Cornell University conducted a study that reviewed the top 10 federal funding agencies and their data management regulations [13]. The study addressed specific criteria regarding data management plans and analyzed the current data management policies at the funding agencies. A key finding of the study is that there is no uniformity in the regulations across agencies. The remainder of this section briefly outlines the existing data management requirements for the major federal funding agencies, as well as select foundations and provides references for the specific policies.

3.1 National Science Foundation

3.1.1 Data Management Plan Requirement

A two-page data management plan must be submitted with all grant applications to the NSF. This plan must detail how the applicant will adhere to the NSF data management and sharing policy. The plan may include the types of data, the standards to be used for data format and content, the policies for access and sharing the data, the policies for reuse and redistribution of the data, and the plans for archiving data [6].

3.1.2 Data Sharing Requirement

Investigators are expected to prepare all significant findings from research done with NSF grants and submit them for publication. This should be at no more than incremental costs and within a reasonable time. The publication should include primary data, samples, physical collections, and other supporting material collected during an NSF-funded project [14].

Steps should be taken to avoid releasing protected or confidential information.

It is the opinion of the NSF that grantees share software and inventions created with NSF funds, though grantees may retain their legal rights to intellectual property. This does not reduce the researcher's responsibility to the scientific community regarding data sharing [14].

Collaborative projects that have subawards are considered one proposal and should only have one combined stat management plan [6]. Proposals for supplementary funding do not require a new data management plan. If an applicant feels a data management plan is not necessary, he or she must include justification in the application [6].

3.1.3 Specific Program Guidance

3.1.3.1 BIOLOGICAL SCIENCES DIRECTORATE

With the increase in the amount of collaboration between the sciences, the Biological Sciences Directorate (BIO) recognizes that the biological subdisciplines may have their own data management standards. Consequently, BIO states that the data management plan should reflect the standards of the area of research being proposed [15].

It is requested that the data management plan be organized according to the guidelines. This includes leading with a description of the data that will be collected along with the standards. Next, a description of what physical and/or cyber resources and facilities will be used to store the data until the end of the project. Third, a description of the media and dissemination methods that will be used to make the data available after the project concludes. Next, the policies for data sharing and public access should be included. Finally, a description of the roles and responsibilities of the parties involved in the grant to the management of the data after the project is completed.

3.1.3.2 ENGINEERING DIRECTORATE

The Engineering Directorate of the NSF also offers additional guidance on creating data management plans. The guidance document is available through the NSF Website and it provides a solid template for researchers to follow [16].

3.1.3.3 EDUCATION AND HUMAN RESOURCES DIRECTORATE

The Education and Human Resources Directorate stresses data on projects involving human subjects [17]. These data should be made available to the public, though subject to constraints imposed by institutional review board (IRB) decisions. The Education and Human Resources Directorate also outlines specific examples for investigators to follow. The examples include a proposal for a workshop that will result in a workshop report, a proposal for developing a new undergraduate course, and a proposal in an education program that requires all projects to report on the graduation rate of participants [17].

3.1.3.4 COMPUTER AND INFORMATION SCIENCES AND ENGINEERING

The Computer and Information Sciences and Engineering (CISE) Directorate is aware of the need to provide flexibility in assessment of data management plans [18]. Each community within CISE has its own definition of data. This will be taken into consideration when reviewing the data management plans submitted by these communities.

3.1.3.5 DIRECTORATE FOR GEOSCIENCES

The Geosciences Directorate recognizes that data management plans are complex. A task force has been created to determine the issues and answers surrounding data management [19]. The Geosciences Directorate has made a request for feedback from the community regarding the new requirements [19].

3.1.3.6 SOCIAL, BEHAVIORAL, AND ECONOMIC SCIENCE DIRECTORATE

This area of the NSF has outlined data management plan requirements [20]. It provides a thorough study on the background of data management plans and describes the required content. The content should include the types of data, how data will be managed, factors that might impinge on the investigator's ability to manage data, the lowest level of data that investigators might share with the scientific community, the mechanism for sharing data, and other types of information that should be maintained regarding the data.

3.1.3.7 MATHEMATICAL AND PHYSICAL SCIENCES DIRECTORATE

The Mathematical and Physical Sciences Directorate division of the NSF gives guidance on creating a data management plan created to each of its disciplines [21]. The areas of astronomical sciences, chemistry, material research, mathematical sciences, and physics all have guidance documentation for investigators to follow when creating their data management plans.

1. Physics states that there is no recommendation of a specific archiving approach [22].
2. Mathematics states that the statement of no data management plan is necessary will suffice for most proposals, provided that the justification is present [23].
3. The Division of Materials Research does not have a recommendation of a specific repository [24].
4. Chemistry recommends the use of a public database. Examples that are provided include The Protein Data Bank, Cambridge Crystallographic Data Centre, PubChem, and the National Institute of Standards and Technology Chemistry WebBook, to name a few [25].
5. Astronomical sciences does not suggest any specific repositories, but requests that these details be provided in all data management plans [26].

3.2 National Institute of Health

3.2.1 Public Access Policy

The goal of the NIH Public Access Policy is to make sure that the public has access to all data generated with NIH funds [27]. The NIH Public Access Policy requires researchers to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive PubMed Central immediately upon acceptance for publication.

The policy applies to any paper that is peer-reviewed, accepted for publication in a journal, and arises from one of the following four circumstances: direct funding from an NIH grant or cooperative agreement active in the fiscal year 2008 or beyond; any direct funding from an NIH contract sign on or after April 7, 2008; any direct funding from the NIH Intramural Program; or an NIH employee.

It is important for an investigator to take this policy into account when publishing a paper. The copyright agreement must ensure that the paper is able to be published in PubMed Central. There are no exceptions [27].

3.2.2 Data Sharing Policy

The NIH policy on data sharing requires the sharing of final research data. It is applicable to applications seeking $500,000 or more in direct costs in any single year of the project period or as a special requirement of a Funding Opportunity Announcement [28].

Applicants must include a data sharing plan with their application. The data sharing plan describes what data will be shared. This should include both final research data and metadata with descriptors. Information on who will have access to the data should also be provided. The data should be shared broadly, and only be limited by laws, regulations, rules, and policies that prohibit the release of specific data. If the data sharing is limited, a rationale must be included in the plan. The location of the shared data should also be included. The ideal place to share data is a data repository with common standards and an established infrastructure.

Data should be made available as soon as possible and for as long as possible. A schedule should be provided including whether any data will be made available prior to publication. The applicant must also include information relating to how the data will be located and accessed. Other researchers must be able to identify the location of any relevant data with ease. It is also necessary that accessing the data is a simple process.

3.3 US Department of Energy

3.3.1 Chief Information Officer

Previously, the chief information officer of the Department of Energy was responsible for ensuring that data was acquired and managed consistent with the policies of the department. This was a broad statement that provided some guidance to researchers with Department of Energy funding. As of October 2013, the Department of Energy's stance on data management became more specific [29].

3.3.2 Office of Science

As of October 2013, all proposals submitted to the Department of Energy Office of Science must include a data management plan. This modification responded to the White House

Office of Science and Technology's revised policy on expanding public access to the results of federally funded research.

The Office of Science has a Statement on Digital Data Management. The focus of the plan is on digital research data and stresses the need for sharing and preservation.

All applications must include a data management plan. It should be no more than two pages and should detail how the data will be shared and preserved. If the data cannot be shared or preserved a rationale must be presented.

The data management plan must describe the method the primary investigator (PI) intends to use to ensure that all data used in published research is available to the public by the time of publication. If any data are used in charts, figures, images, etc., they must be made available to the public. This can be done by including the data as a supplement to the article or by providing it in a repository. The published article must tell the reader where the data can be accessed [30].

If the researcher is planning to work at an Office of Science User Facility, he or she must consult the published data policy of that facility and reference the policy in her data management plan.

3.3.3 Climate and Environmental Sciences Division

The Department of Energy requests that data of potentially broad use in climate change research be archived [31]. This is intended to promote networking among the members of the global climate-change community and inform preparation of technical and informational reports, and sponsorship of scientific conferences.

3.3.4 North American Research Strategy for Tropospheric Ozone Projects

If a researcher is applying for a grant under the North American Research Strategy for Tropospheric Ozone (NARSTO) initiative of the Department of Energy, a data management plan must be submitted [32]. A guidance document is available instructing an applicant how to tailor their plan in the NARSTO context. More than 20 sample data management plans are provided depending on the type of funding one is applying for. The samples range from organization focus, data and metadata reporting, data documentation and archiving, and data systems management.

3.4 US Department of Education

3.4.1 Institute of Education Sciences

Researchers with funding provided through the National Center for Education Research and the National Center for Special Education Research, the two research centers of the Institute for Education Science (IES), are required to include data sharing plans [33]. The institute believes that sharing data will lead to more rigorous peer review, and so the production of more robust findings.

This requirement was instituted for 2013 grants and continues. All plans must describe the data to be shared, the method of sharing, the documentation that will accompany the data, the plan for keeping personally identifiable data confidential, the projected timeline for sharing data, the roles of the staff in the management and retention of the data, and the cost for sharing the data. The two methods of sharing described are through the Principal Investigator (PI) directly or a data archive/repository.

Data are expected to be made available upon publication of the attendant findings. The IES recognizes that this may not occur until the research is complete. The IES permits budgeting of data sharing costs. These costs may include preparing the data to be archived, as well as the data documentation. A guideline with further information is provided at the IES Website [33].

3.5 US Environmental Protection Agency

The US Environmental Protection Agency (USEPA) does not have an official data management and sharing policy. However, it does supply several documents describing the importance of data management that should be taken into consideration by researchers with USEPA funding.

3.5.1 Open Government Data Quality Plan

The Open Government Data Quality Plan is an internal policy of the USEPA [34]. It outlines the implementation of data quality framework stressing open communication and monitoring between the government and funded teams. It also discusses steps for reporting data and access to budgets regarding funding.

3.5.2 Survey of the Environmental Protection Agency and Other Federal Agency Scientific Data Management Policies and Guidance

In 2010, the USEPA released a report of its data management practices in comparison to other funding agencies [35]. The report is very detailed and covers data management plans in relation to assets or liabilities, the full data lifecycle, ability to identify metadata, and other data management related topics. The report concludes that there is yet to be a comprehensive data management policy among federal agencies, demonstrating that Data Management requirements continue to evolve.

3.6 US Agency for International Development

3.6.1 Data Quality Assessment

The five data quality standards the US Agency for International Development (USAID) holds key are validity, reliability, precision, integrity, and timeliness. The USAID encourages data quality assessment to ensure that data generated from USAID-funded research is up to their standards. There is no specific way to conduct a data quality assessment, stating a memo in the file or a formal report may be necessary, but the USAID suggests one be done at least every three years [36].

3.6.2 American Schools and Hospitals Abroad

The USAID provides funding under a number of different grant instruments. The award under the American Schools and Hospitals Abroad program requires a monitoring and evaluation plan [37]. Applicant must specifically outline how, when, and by whom data will be gathered, analyzed, and used. Although this plan may be brief in the application stage, it could lead to a larger plan.

3.7 National Oceanographic and Atmospheric Administration

3.7.1 Data Sharing Policy

Data sharing is one of the many facets of data management. The National Oceanographic and Atmospheric Administration (NOAA) has separate requirements for data sharing [38]. The NOAA is concerned with making the data visible, accessible, and understandable. Data sharing is limited by privacy concerns and any other superseding laws or regulations.

Under the NOAA data sharing policy, all data using NOAA funding must be shared in a timely manner. Typically, this is interpreted to mean no later than two years after the data are collected or created. A two-page data sharing plan is required as part of the project narrative. The information supplied should include the types of environmental data created, the standards to be used to format the data, policies accessing data preservation, previous data sharing experience, and procedures for providing access, sharing, and security. Data sharing plan templates are provided by the NOAA along with case examples.

When deciding on where to share the data, a researcher should first consider NOAA facilities that archive data. The NOAA Procedure for Scientific Records Appraisal and Archive Approval describes the process of contacting an NOAA archive [39]. It should be noted that failure to publish the data may result in diminished awards and denial of any future awards.

3.7.2 Oceanographic Data

When oceanographic data and related information are acquired with federal funds, the data must be submitted to the National Oceanographic Data Center [38]. This is a national repository that holds some international data as well. The National Oceanographic Data Center has acquisition specialists that will assist researchers with storing their data.

3.8 National Aeronautics and Space Administration

3.8.1 Data Rights and Related Issues

The National Aeronautics and Space Administration (NASA) details its data management requirements in its Data Rights and Related Issues policy [40]. NASA promotes the full and open sharing of all data. It lists the organizations with which data should be shared, such as academia and the general public. Archived mission project data management plans are available to the public. Although its official Data Management stance is general, specific areas of NASA funding require more information.

3.8.2 Earth Science Missions

Data management plans are required under the Earth Sciences division of NASA [41]. Guidance for creating a plan is provided by the agency. The Data management plan must include information on the development, maintenance, and management responsibility, the change control (plans for modification and updates to this document over time), all relevant documents, the project objectives, science objectives, mission summary, instrument overview, mission operations, science operations, post-mission stewardship and access, science data product summary, associated archive products, special considerations, and a section on acronyms.

3.8.3 Heliophysics

Heliophysics research provides its own requirements for Data management plans [42]. Guidelines are available for creating Data management plans, which must include preparing, accessing, using, and archiving heliophysics data. This policy stresses NASA's overall open access data policy mentioned above.

3.9 National Endowment for the Humanities

3.9.1 Office of Digital Humanities

All applicants must submit a data management plan supplementing their grant application [43]. The plan should be no more than two pages and answer two questions: What data are generated by your research and what is your plan for managing the data? The plan should articulate, in a clear and concise manner, how sharing of the primary data will be accomplished. The rights and responsibilities of the parties involved in relation to management and retention of data must also be outlined. Any costs of data sharing should be included in the budget. The matters that should be addressed are the type of data, how they will be managed and maintained until sharing is possible, factors that will impede sharing of data, the lowest level of data that can be shared, the mechanism for sharing data, and analytical and procedural information regarding the data. The data are expected to have a timely and rapid distribution; however, no minimum time requirements are provided.

3.10 Institute of Museum and Library Services

3.10.1 Open Government

The Institute of Museum and Library Services (IMLS) does not have specific data management requirements. It does stress open access government. It is a goal of IMLS to make the agencies work more transparent and encourage public participation. Meeting this goal may fall under data sharing guidelines.

3.11 American Heart Association

3.11.1 Guide for Affiliate Research Awards

The American Heart Association provides little guidance on data management. However, in its guide for affiliate research awards, it specifically mentions research misconduct including data falsification [44]. Implementing a strong data management plan can help a researcher defend against claims of scientific misconduct. In general, researchers should become familiar with creating data management plans regardless of the requirements because it is likely more and more agencies will look for them in the future.

3.12 Alfred P. Sloan Foundation

The Alfred P. Sloan Foundation offers only general guidance on the necessity of data management [45]. It makes reference to the importance of handling a wide variety of digital

data and encourages researchers to make data publicly accessible. As the trend from general guidelines to specific requirements continues, researchers should create data management plans that follow these objectives.

3.13 Centers for Disease Control and Prevention

3.13.1 Data Sharing and Release Policy

The Centers for Disease Control and Prevention (CDC) policy on data sharing includes all data collected for the CDC by institutions. The policy specifically mentions data generated through grants, contracts, and cooperative agreements. The policy requires that all data are released or shared as soon as possible without disregarding privacy and confidentiality concerns or other applicable laws [46]. The policy is vague and does not require a specific data management plan, but every researcher funded by the CDC should expect to release and share data in a reasonable timeframe.

3.14 Department of Defense

3.14.1 Department of Defense Directive 3200.12

Similar to the CDC, the Department of Defense (DoD) only gives vague instructions regarding data management. It requests that researchers "establish and maintain a coordinated and comprehensive program to document the results and outcome of DoD-sponsored and/or performed research and engineering and studies efforts and provide access to those efforts in an effective manner consistent with the DoD Mission."

There is no requirement to submit a data management plan with grant proposals; however, it is expected the researcher have one. Many DoD grants come with stipulations and confidentiality agreements. Researchers must take these clauses into account before sharing any data to make sure they are in compliance.

3.15 National Institute of Justice

A brief data archiving strategy must be included in any application for National Institute of Justice (NIJ) research grants [47]. This should describe how the data will be prepared and documented. The NIJ gives details on its requirements and requires that the data archiving strategy includes information regarding the data formats, which software will be used to archive the data, a description of the procedures for data collection, and confidentiality details in the research involves human subjects.

Most NIJ-funded research must have its data sets archived with the National Archive of Criminal Justice. The data must be submitted 90 days before the end of the project period.

3.16 National Institute of Standards and Technology

The National Institute of Standards and Technology does not have requirements directed solely toward data management. In its document addressing information quality standards and administrative mechanism, data release is mentioned. It is required that data from

National Institute of Standards and Technology grants is disseminated for public use or made available through an ad hoc request that results in the steward no longer controlling the data [48].

3.17 US Department of Agriculture

The US Department of Agriculture Cooperative State Research, Education, and Service require all funded research be submitted into the public domain without restriction. There are few additional instructions provided to researchers. When there is a patent application, exceptions are sometimes allowed [49].

3.18 Office of Research Integrity

The Office of Research Integrity (ORI) is not a grant awarding body. It serves as a regulatory committee regarding misconduct and provides information and training regarding the RCR, among other things. Data management is one of the main areas of focus within RCR and the ORI offers a variety of information on the subject [11].

3.18.1 Guidelines for Responsible Data Management in Scientific Research

A course regarding data management is offered through ORI entitled Guidelines for Responsible Data Management in Scientific Research [50]. A 35-page course outline covers general data management, data ownership and retention, data collection and record keeping, data storage and protection, data sharing and publication, human subjects research, animal research, and research team leadership and communication. Case studies and real-world examples are also provided in this training. It is highly recommended every researcher take the time to review this document.

3.18.2 Data Acquisition and Management

ORI also offers training in data acquisition and management [50]. This exercise is intended for individuals with some knowledge of data management. The focus is on two separate case studies, what to look for, and the outcomes. It is the responsibility of the PI to offer continuous training in RCR to the research team and students. This training is a good way to keep data management issues present in researchers' minds.

3.18.3 Educating Clinical Research Staff in Clinical Data Collection and Data Management

This is an interactive tutorial provided by the ORI in conjunction with Saint Jude's Children's Research Hospitals and Cure4Kids [50]. The focus of this training is intended to address the needs of clinical staff members and covers issues regarding patient collection and recording. Data management plans are not limited to clinical research.

3.18.4 Interactive Data Management Module

ORI provides a link to Northern Illinois University's data management module [50]. This is an interactive training geared toward researchers with a very basic familiarity with data

management. All areas of data management are covered in this tutorial and it can take up to 4 h to complete. One does not need to finish the entire course in one session. Although a PI can benefit from taking this module, it is a good practice for the PI to require training similar to this for his or her research staff.

4. KEY REGULATORY MANDATES

To date, the key regulatory mandates regarding data management are those promulgated by the NSF and NIH and subsequently modified and adopted in whole or in part by other federal agencies and private funding sources. The pillars of these mandates are: (1) that data be collected and stored in a manner that permits the scientific community and interested others to access, review, and make further use of that data. This mandate recognizes both the value (intrinsic and extrinsic) of that data collected and the need to safeguard that data for future use and the dramatic increase in problem spaces that require that data and scientific information be made easily and cheaply available to scientists from their own and other disciplines. (2) That the findings of scientific research, especially those resulting from publicly funded research projects, be made publicly available as widely and as quickly as practicable. Again, this mandate arises from the recognition of the need for scientists across distances and disciplines to have access to the latest scientific information in a timely manner to foster the cross-pollination of scientific thought and to ensure that scientific findings are further tested through replication and extension in order to solidify their basis as reliable knowledge.

The NSF data management policy simply states that "[I]nvestigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants."

The NIH simply stipulates that: "Investigators seeking $500,000 or more in direct costs in any year should include a description of how final research data will be shared, or explain why data sharing is not possible" [51]. It then goes on to provide the following detail regarding the contents of the plan and opportunities for covering the costs of the plan within the budget justification:

> Applicants who are planning to share data may wish to describe briefly the expected schedule for data sharing, the format of the final dataset, the documentation to be provided, whether or not any analytic tools also will be provided, whether or not a data-sharing agreement will be required and, if so, a brief description of such an agreement (including the criteria for deciding who can receive the data and whether or not any conditions will be placed on their use), and the mode of data sharing (e.g., under their own auspices by mailing a disk or posting data on their institutional or personal website, through a data archive or enclave). Investigators choosing to share under their own auspices may wish to enter into a data-sharing agreement…
>
> Investigators working with archives can get help with data preparation and cost estimation. Investigators who are concerned about paying for data-sharing costs at the end of their grant can make prior arrangements with archives. Investigators facing considerable delays in the preparation of the final dataset for sharing should consult with the NIH program about how to manage this situation, such as requesting a no-cost extension.

These mandates, and similar language provided by other federal funding sources, address data management simply by requiring plans to be included in requests for funding.

However, to date, the plans do not play a significant role in the peer review process. In this sense, they function as an institutional assurance and no more. A regimen of stronger mandates would require not only explicit attention during peer review, but also more explicit recognition of the need to fund data management activities and, eventually, the need for the time and effort required to develop, populate, and monitor data management systems within the context of a research project to be recognized by investigator's departments as activity worthy of measurement for the purposes of tenure, promotion, and merit pay increases.

However, this picture may change as agencies respond to a White House executive order on increasing access to federally funded scientific research. The order requires agencies that fund in excess of $100 million in research to develop concrete plans for data management and open access.

In terms of digital scientific data and publications, the agency plans must contain the following elements:

1. a strategy for leveraging existing archives, where appropriate, and fostering public–private partnerships with scientific journals relevant to the agency's research;
2. a strategy for improving the public's ability to locate and access digital data resulting from federally funded scientific research;
3. an approach for optimizing search, archival, and dissemination features that encourages innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research;
4. a plan for notifying awardees and other federally funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, regulatory changes);
5. an agency strategy for measuring and, as necessary, enforcing compliance with its plan;
6. identification of resources within the existing agency budget to implement the plan;
7. a timeline for implementation; and
8. identification of any special circumstances that prevent the agency from meeting any of the objectives set out in this memorandum, in whole or in part.

With respect to public access to scientific publications, the order requires that each agency shall:

1. Ensure that the public can read, download, and analyze in digital form final peer-reviewed manuscripts or final published documents within a timeframe that is appropriate for each type of research conducted or sponsored by the agency. Specifically, each agency:
   a. shall use a 12-month postpublication embargo period as a guideline for making research papers publicly available; however, an agency may tailor its plan as necessary to address the objectives articulated in this memorandum as well as the challenges and public interests that are unique to each field and mission combination, and
   b. shall also provide a mechanism for stakeholders to petition for changing the embargo period for a specific field by presenting evidence demonstrating that the plan would be inconsistent with the objectives articulated in this memorandum.
2. Facilitate easy public search, analysis of, and access to peer-reviewed scholarly publications directly arising from research funded by the federal government;
3. Ensure full public access to publications' metadata without charge upon first publication in a data format that ensures interoperability with current and future search technology.

Where possible, the metadata should provide a link to the location where the full text and associated supplemental materials will be made available after the embargo period;

4. Encourage public–private collaboration to:
   a. maximize the potential for interoperability between public and private platforms and creative reuse to enhance value to all stakeholders,
   b. avoid unnecessary duplication of existing mechanisms,
   c. maximize the impact of the federal research investment, and
   d. otherwise assist with implementation of the agency plan.
5. Ensure that attribution to authors, journals, and original publishers is maintained; and
6. Ensure that publications and metadata are stored in an archival solution that:
   a. provides for long-term preservation and access to the content without charge,
   b. uses standards, widely available and, to the extent possible, nonproprietary archival formats for text and associated content (e.g., images, video, supporting data),
   c. provides access for persons with disabilities consistent with Section 508 of the Rehabilitation Act of 1973, and
   d. enables integration and interoperability with other federal public access archival solutions and other appropriate archives.

Regarding agencies' plans to provide public access to scientific data in digital formats, the order stipulates that each agency will:

1. Maximize access, by the general public and without charge, to digitally formatted scientific data created with federal funds, while:
   a. protecting confidentiality and personal privacy,
   b. recognizing proprietary interests, business confidential information, and intellectual property rights and avoiding significant negative impact on intellectual property rights, innovation, and US competitiveness, and
   c. preserving the balance between the relative value of long-term preservation and access and the associated cost and administrative burden.
2. Ensure that all extramural researchers receiving federal grants and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified;
3. Allow the inclusion of appropriate costs for data management and access in proposals for federal funding for scientific research;
4. Ensure appropriate evaluation of the merits of submitted data management plans;
5. Include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies;
6. Promote the deposit of data in publicly accessible databases, wherever appropriate and available;
7. Encourage cooperation with the private sector to improve data access and compatibility, including through the formation of public–private partnerships with foundations and other research funding organizations;
8. Develop approaches for identifying and providing appropriate attribution to scientific data sets that are made available under the plan;

9. In coordination with other agencies and the private sector, support training, education, and workforce development related to scientific data management, analysis, storage, preservation, and stewardship; and
10. Provide for the assessment of long-term needs for the preservation of scientific data in fields that the agency supports and outline options for developing and sustaining repositories for scientific data in digital formats, taking into account the efforts of public and private sector entities.

Agency responses to this order will address some of the concerns raised here, for example, they will at least indicate that data management and public access plans must be made part of the peer review process, but motivation and incentive for faculty and research institutions to change their practices and their culture in this regard remain distant prospects.

5. KEY PERSONNEL AND UNIVERSITY COMMITTEES DESIGNATED TO IMPLEMENT REGULATORY MANDATES

Universities address the wide range of issues related to data management from a number of organizational directions. Libraries commonly have staff members, departments, and committees intended to support faculty, staff members, and students meet data management requirements and access library-based data management services. They often also cover issues related to copyright, rules for metadata use, connection with international and discipline-based repositories, and requirements related to data sharing and open access publishing. Research and Sponsored Programs offices may have staff members who can assist faculty in developing appropriate data management plans or accessing tools and templates for completing their plans (e.g., DMPTools [52]). Information Technology Services departments often have staff members and committees that address issues of data storage and transfer. Issues related to data management are also often addressed by legal counsel, Provost's Office committees, research committees, and arise as matters addressed by faculty senates. Data management is also one of the core elements of what has come to be termed RCR and so the development of policies and procedures often falls to RCR committees. Not all institutions use all of these institutional options or use them in the same ways; nor are all of them necessarily charged with ensuring compliance with regulatory mandates, but their goals are generally aligned around ensuring that legal, ethical, regulatory mandates are met in ways that support access to and the sharing of research data and information. Generally, the university units and committees charged with ensuring regulatory compliance for data management include the following.

5.1 Principal Investigators

In terms of project implementation and the ways in which specific agency requirements regarding data management and publication of finding are met and managed, the responsibility falls to the PIs as the leaders of the research project and the signatory on the grant award. They set the tone and assign the responsibilities to their staff and collaborators. They will plan out the project and have the final decision on all matters regarding how data are

collected, stored, managed, and shared. The publication of the final results is a central responsibility of the PI as well, and so they are responsible for meeting all requirements regarding publication and access to results. The PI is also required to meet any university requirements regarding data management and sharing and to abide by university policies as they relate to data security, confidentiality, data ownership, intellectual property, and copyright.

5.2 Research Office

The Research Office generally houses a number of units that play important roles in developing and implementing data management policies and in ensuring adherence to regulatory mandates.

5.3 Research Compliance

Research compliance often promulgates general university policy as it relates to the ownership and use of research-related data, and to the requirements for meeting federal mandates (see Section 6). The Research Compliance Office is often also the home of the unit that addresses research misconduct, which may also involve issues related to data management and reporting as required by federal agencies (see Section 6).

5.4 Responsible Conduct of Research Committee

In response to NIH and NSF requirements that students participating on grants funded by those institutions receive training in the RCR, many institutions have RCR committees, which are usually housed within the university research office and generally made up of faculty and administrative staff representing those departments most commonly affected by the NIH and NSF mandates. Generally, these committees are responsible for a number of items, including design and implementation of core RCR programs and coordination with department-specific RCR training programs. Data management is one of the core elements of RCR training. Therefore, although RCR committees may develop RCR standards and implement policies that affect RCR, they are unlikely to make policy specific to data management, but simply ensure that students and faculty are properly apprised on issues related to data management through RCR training.

5.5 Sponsored Projects

This is the office that will assist researchers with applying for funding. In this capacity, it often assists investigators in developing appropriate data management plans based on the guidance provided by the funding agency or by directing them to specific data management tools. It is also responsible for reviewing proposals before they are submitted to ensure that all federal, state, and institutional requirement are met. For proposals to federal funding agencies, this includes a long list of assurances and, where mandated, inclusion of the required data management plan.

5.6 Institutional Review Board

The IRB is a committee responsible for approving, monitoring, and reviewing biomedical and behavioral research involving human subjects. Often these proposals come with

confidentiality restrictions regarding information provided by or related to the human subjects. The IRB can be instrumental in determining the confidentiality concerns and offering guidance and regulations on how various types of data can be held and for how long.

5.7 University Libraries

University librarians, staff members, and committees are often experts in the regulations, technologies, services, and discipline-specific requirements related to data management and providing access to published material. They may be usefully consulted on matters relating to storage, archiving, curation, transfer, and the appropriate use of metadata. Libraries often house the university data repository, which is often available free of charge to investigators on campus. Storing data in a repository is often a safe way to ensure the integrity of the data and fulfill funding agency requirements.

5.8 Outside Parties

5.8.1 Journals

At present, journal involvement with research data most often arises in connection with disputes over findings or cases involving potential misconduct. In this role, many journals have become significant players in areas of data management related to data integrity. However, it is becoming more common for journals to request that data be made available along with the publication of findings. As this practice becomes more common, journals will play an ever larger role in data management. Discussions of open access writ large are beyond the purview of this chapter; however, as federal mandates stressing access evolve, the role of journals in this area will become ever more central.

5.8.2 Professional Societies

Professional societies often provide disciplinary standards and best practices guidance to their communities. They may also provide specific guidance on how to prepare appropriate data management plans. In some cases, they may also recommend data archives.

6. COMMON COMPLIANCE CHALLENGES

Every project will have unique situations that arise and must be dealt with on a case-by-case basis. However, the ORI in the Department of Health and Human Services (HHS) has recognized frequent issues that arise involving data management. These are summarized here and can be found in more detail at the ORI Introduction to the RCR Guide and the ORI Data Management Guidelines.

6.1 Creating Internal Policies

Many universities lack formal policies and tend to rely on “industry standards.” Although a large number of funding agencies require some form of data management and record

keeping, it is still beneficial for the university to have its own internal policies. In the event an ownership or access issue arises, universities without policies could find themselves in very difficult places. Because data management is a pillar of the RCR, any data management policy should supplement existing RCR and research misconduct policies and not contradict them.

6.1.1 University-Wide Policy

Different disciplines or departments may have their own particular policies, but it is important to have a university-wide policy regarding data management practices and procedures. These policies should protect the university and the researcher. It is difficult to regulate all the aspects of data management, but data usage, ownership, retention, and access should be addressed.

6.1.2 Data Usage

Internal policies should cover data usage. It may be helpful to identify a global standard, such as the ORI language, because this policy may cover a wide variety of disciplines. Any policy should grant the Investigators appropriate use of the data. The policy should stipulate that all data usage must be in accordance with contractual commitments by the university. The policy may also encourage investigators to create written plans regarding ownership, rights, access, and permissions for data use at the outset of projects.

6.1.3 Custody of Research Data

Because research awards from funding sources are usually granted to the institution, the university is generally the owner of any research data collected under the auspices of the university (i.e., by faculty in the normal course of their duties). The custody and stewardship of the research data are usually in the hands of the PI. In some cases, research participants or governments with jurisdiction of property from which samples are taken may seek to retain forms of ownership or rights to data collected from their samples. In such cases, it is important to document or address the apportionment of those rights in an informed consent document or some other form of written agreement.

Some university policies further expand upon this, explaining the chain of custody of the data if the PI becomes unavailable. This should include what will happen to the data in the event of research misconduct proceedings. The research misconduct policy and the data management policy should not contradict each other and ensure the university has access to the data.

As custodian of the data, the PI should also be charged with maintaining the integrity of the data. This could include responsibilities toward data collection, analysis, or storage among others.

6.1.4 Data Retention

Although many funding agencies have their own data retention policies, it may be in the interests of the university to implement their own standards. In some cases, there may be benefits to retaining data, because as the owner of the data, the university can set time limits the custodian must abide by. However, there are also cases where sensitive or personal data, such as video or audio need to be retained. In such cases, it is important to balance the needs of the investigators, and often the policies of the journals to which they submit articles, and

the research participants, whose interest is in seeing that the data are destroyed as soon as is practicable. Whatever limits are set, the clock usually starts after submission of the final research data.

It is advised that alternative retention requirements are given if the work belongs to a student. These may include retaining the data until the student graduates or it is clear that the work has been abandoned.

Data retention policies need not limit themselves to funded projects. It is important to define who the policy applies to and possibly include provisions that exclude unfunded work for certain sections.

6.1.5 Data Access

Internal university data management policies should state that the university has access to all research data. Universities should not limit themselves to data regarding research preformed at the university and should include any research supported by university administered funds.

It is important that researchers not feel alienated by these clauses and be given ample warning before the university requests the data absent an emergency.

6.1.6 Confidentiality

Data management policies should ensure that all confidentiality agreements and requirements put forth in human subjects protocols and informed consent documents regarding the confidentiality of data be honored. Federal funding often requires that specific procedures be followed in order to safeguard confidential or otherwise sensitive data. Policies should specify adherence to such procedures.

6.1.7 Data Sharing

As mentioned previously data sharing is seen as a tenet of the scientific community. More and more organizations encourage data sharing. Striving for data sharing is something a university must consider, but with the demand for sharing increasing including a data sharing clause can be expected from faculty and administration alike.

It is important that confidentiality concerns are taken into consideration with data sharing. When a patent or other copyrighted material is involved researchers may be reluctant to share their data. Researchers should be made aware of any exceptions to data sharing requirements.

6.1.8 Data Transfer

If an investigator separates from his or her institution, he or she may wish to take the research to the new organization. The university is the owner of the data and has the right to deny that request; however, most organizations allow for some form of agreement. Mentioning this in the policy and even putting a blanket statement this will be handled on a case-by-case basis can make the transition easier for separating investigators. Make sure that there are people in place to handle the transfer.

Some of the transfer arrangements may include the university keeping the originals, whereas the investigator takes copies or vice versa. There can also be steps put into place where the investigator can take the research, but the university retains the rights to access all the data.

6.1.9 Data Ownership

Prevailing higher education practice indicates that the university is the owner of all the research data. This should be made clear throughout the policy. This statement should take preference over any other policy that mentions this topic, whether they be intellectual property policies or human resource issues.

6.1.10 Disputes

The policy should include a reference to where to handle disputed issues. This could be handled by going to the vice president for research or the RCR Committee for example. The details do not have to be written out, but just some basic direction where the individual should go if they have any issues regarding the internal data management policy.

6.2 Project-Based Data Management Plans

Many funding agencies require data management plans with their grant applications. Several of these plans are limited to two pages. Some organizations, such as NASA, provide sample data management plans for applicants. It is a good idea for researchers to examine these plans before creating their own.

Universities will normally have an administrator who will be able to assist researchers with drafting their project-based data management plans; there are also a number of resources available online. The California Digital Library has a free data management plan tool for investigators at https://dmp.cdlib.org/. This module is tailored for each funding agency. This can be a great help to ensure all the requirements of the funding agency have been met.

Even if the specific funding agency does not require a data management plan it is advised that the PI create one for the team. Data management plans organize the research, guide the team, and streamline the process. The following should be considered when the PI creates the data management plan.

6.3 Project Needs

To avoid confusion and make the project run smoothly, the PI should outline the needs of the research project. These should cover all the areas of data management from data collection to data reporting. Breaking them down step by step will give a clearer picture of what needs to be done regarding the data.

6.3.1 Team Member's Skills

As researchers work together longer, they will become aware of new or increasing skills, as well as areas where individuals need improvement. Most universities offer a variety of training options and the PI should be aware of what the people on the team can handle and where they can get more assistance.

6.3.2 Roles and Responsibilities

If each member of the team knows what is expected of them, the research can run smoothly. This not only benefits the PI because it save the project time, but it will also ensure the integrity of the project because the team will be aware of expectations and what to perform.

6.3.3 Potential Problems

If the researcher is aware of potential problems before they arise, he or she will also be aware of potential solutions. It is best to evaluate the options available, especially if a data ownership or confidentiality issue may arise. That way the PI will know what steps to take next.

6.3.4 Timespan

Data management plans can keep the project moving forward. The PI should look at the finished plan and set goals, including specific dates, and make sure that the team is aware of these aims. The goals should also include any mandatory deadlines set by the funding agency. Progress should be checked against the timeline to ensure that the project will be finished in a timely manner.

It is recommended that the timeline extends beyond the date of publication, especially if there is a data sharing requirement. Many organizations give the PI a time limit before the team must share its data. This should be noted and honored or else the PI risks violating the grant agreement.

6.4 Types of Data

Most data management requirements are less than a decade old. This is a changing landscape and some researchers may be reluctant to change. In 2012, the journal article "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements" was published. The authors surveyed current and prospective NSF PIs from Cornell regarding data management plans. The responses to the survey indicated that 62% would be interested in guidance for writing a data management plan.

The results of the study showed two challenges when offering support for data management plans. The first challenge was that researchers were confused regarding the definition of what is considered data. Universities must be prepared to provide guidance on this issue. It is important that a set definition for data is provided and that researchers understand it. This can be accomplished with training or through guidance provided by data management plan servicers.

The second issue was the vast array of digital content. As technology continues to advance, it is important that the people giving guidance on data management plans are aware of what storage and analytical tools are available to address the specific needs generated by various kinds of content and the ways in which that content is most effectively accessed.

6.5 Standards for Data and Metadata

6.5.1 Understandings

Metadata is a set of data that gives information about other data. Standards vary among disciplines. Any persons giving data management advice must know what field they are working with before attempting to answer questions dealing with standards. The same study mentioned previously [53] showed not all researchers are aware of whether their data or metadata conforms to the standards of their discipline.

There was also an apparent lack of understanding regarding the definition of metadata, along with a reluctance to use metadata in one's research.

6.5.2 Recommendations

NSF's recommendation that researchers specify "standard to be used for data and metadata" should be interpreted to mean formally, or de facto, standards of the discipline, not within a single laboratory.

The same confusion appears to exist among data management plan servicers. This should be seen as an opportunity to create and implement training programs. These could consist of online or face to face trainings and explain data standards and how one creates metadata.

6.6 Data Sharing

The dilemmas a research may face when sharing data are the necessity of protecting the privacy of all research participants, maintaining the confidentiality of sensitive information, monitoring and preventing inappropriate use of the data shared [54].

6.6.1 What to Share

When sharing data, two challenges researchers face. The first is determining what the investigator should share. Research can generate a large amount of data, but not all if it may be relevant to the research. This is a difficult question for researchers to answer. They should look at their final discoveries and make sure that the data shared can recreate their research. They should also look to their peers for guidance and see what has worked best for other researchers in the past.

6.6.2 Confidentiality

There is a concern regarding confidentiality and data sharing. A significant amount of research experiments contain confidential information, especially those involving human subjects. Institutions should make sure their research infrastructures are up to date and in place. Educating investigators on confidentiality laws will also ease issues they have with sharing their data.

6.6.3 Where to Share

NIH has internal systems for sharing data, but not every funding institute is the same. Certain publications may also offer a data repository if the article is to appear in one of their editions. Some institutions offer internal repositories that are open to the public for researchers to share their data. These services are generally free and it is recommended they are used by investigators. Some disciplinary repositories exist as well. These repositories include the Inter-University Consortium for Political and Social Research, GenBank, and ArXiv. There are many other options available as well. Finally, a researcher may use public services such as Good Docs or Dropbox.

6.7 Financial Requirements

Many of the data management requirements make mention of costs. They promote data management initiatives at the lowest cost possible to investigators. Even with this caveat, there are some concerns regarding the costs of implementing data management plans.

6.7.1 Storage and Retention

When dealing with large amounts of hard data, the cost of storing the items can add up. Some biological specimens can require refrigeration, which may be more costly than storing laboratory notebooks. The length of retention should be noted as well. Even if the funding agency has a specific requirement, this should be checked against internal university policies. Whichever one is longer will be the one the investigator needs to follow. Costs should be examined before the start of the research and extend through the retention period. If these financial concerns are addressed early, they should not be a problem for the investigator later on.

6.7.2 Digital Data

Research can generate large amount of digital data. The larger amount of data, the more the researcher has to work with—but also the higher the costs. Sometimes analyzing large amounts of digital data will require a third party. The fees these companies charge can vary. It is important the PI knows upfront their plan for analyzing large electronic files.

6.7.3 Protection

Certain data come with confidentiality restrictions or can be sensitive to public knowledge. Regardless of whether the data are physical or digital, protecting these data can come with some costs. If the data are physical, a safe, a locked room, or even video monitoring may be required.

Digital security comes with a variety of options. This can include encrypting files, creating secure logins, monitoring who accesses the files, and using a private server. Many universities have internal technology departments who will be able to assist the PI in protecting the digital data, but in some cases a third party may need to be used. It is good planning to make sure the researcher is aware of these costs.

6.7.4 Destroying Data

The cost of destroying the hard data should also be calculated before beginning the project. This could include shredding services or hard drive wipes. It is difficult to predict the cost of the destruction of data because some retention periods are as far as five years after the project completion, but an estimate for these services should be recognized.

6.8 Communication

Even if the PI has a very detailed data management plan, he or she cannot expect everything to run smoothly for the duration of the project with engaging in communication. The PI needs to communicate well with the team. He or she should personally educate the members regarding RCR issues, encourage team members to engage in open discussion regarding the data, and promote open communication regarding any issues or problems the team might face. Encouraging open communication among the team will make sure that everyone is aware of details involving the grant and helps strengthen the integrity of the data.

6.8.1 Communication Plan

By the time the project commences, the PI should implement a communications plan. It is best if this plan is in writing and available to all members, whether in hand out form, e-mail, or available on a Website. Copies may even be posted in the laboratory.

It is important that the communications plan establish a chain of command. It should also outline who is able to make decisions regarding the research and data. The team members must be aware of what information is to be communicated, who will be told, and how they will be informed. The plan should also let the team know what forms of communications should be in writing, electronic, or oral.

It is a good idea to integrate a communications plan with data collections issues. The approach for collecting data and who to inform about that steps have been taken can be integrated. This should outline the way data collection will be recorded as well and serve as a benchmark among the project's timeline.

A sample idea would be requiring team members to send an e-mail every time they leave the laboratory. The communication should include how much data was collected, who collected the data, in what format was the data collected, and so on. This would create an easy paper trail promoting the integrity of the data.

The communication does not have to be limited to internal dialogs. The plan can include communication within the institution, including updates with the Office of Sponsored Projects. They may also mention communications directly with the funding agency as well as set time to speak with peers working in different laboratories or institutions.

6.9 Leadership

It is necessary that the PI be the leader of the research as well as the research team. To be an effective leader, the PI should conduct himself or herself in a manner that provides the project's goals. The PI needs to set the vision of the project and control the way it is heading. If the PI is able to involve every team member in the goal-setting process, it will unify the team, and the work will demonstrate the care put into the research. Creating goals will also provide the PI with a way to self-check the project and note who is accountable for which parts of the research.

It is necessary that the PI remains a figure of authority but does not alienate team members or appear unapproachable. Open communication should be encouraged or problems may arise that the PI is unaware of, which could affect the data. Fostering trust among the team members will allow the PI to be more aware of what is happening on the project.

An effective leader strengthens the integrity of the data. The PI should make sure he or she is creating a healthy environment for the team and leading by example with specific goals and open communication.

6.10 Managing Conflicts

It is a safe assumption that conflicts will present themselves over the course of a project. These issues may be between team members or a team member and the PI. As the leader, it is the PI's responsibility to recognize and deal with the conflicts. Failure to act quickly can endanger the integrity of the research.

Some issues that could arise include conflicting outlooks between team members, lack of recognition, authorship disputes, feeling over- or underworked, a belief that not all of the resources are being used to their full potential, refusal to follow protocols, frustration with the project, unhappiness over the goals of the project, or refusal to accept negative outcomes.

If the PI has created an environment in which the team members feel like they can openly communicate among each other and express themselves, these issues will be brought to the PI's attention before they get out of control and cause a major issue within the laboratory.

The PI should provide constructive feedback and address the problems as soon as they arise.

One of the best ways to deal with an issue is to simply listen. Active listening, which is done with body language, attentiveness, and awareness of the issues, will let the offended party know he or she is being taken seriously. The PI should never interrupt, react to, or correct the team member. Once the person is finished expressing the initial concerns, the PI should continue in a calm and open manner.

The PI should not show judgment or pick sides, but rather talk softly and explain his or her views on the issues. This might include emphasizing contributions the team has made or reiterating the initial goals of the project. It is important the PI continue to stress his or her approachability and not shut down open communications.

Although the PI cannot predict every conflict that will arise, he or she should address them in a calm and consistent manner. The PI should focus the discussion on the conflict itself and not let any member of the team feel attacked. If a strong communication plan is in place and the PI has demonstrated leadership skills, then the team will come to the PI with any issues. By handling these issues in a quick and consistent manner, the PI is preserving the integrity of the research.

7. ADDRESSING NONCOMPLIANCE

7.1 Research Integrity

Data management policies and regulations are put into place to ensure the integrity of the research. A violation of data management practices and procedures will normally amount to research misconduct.

In 2005, processes were put into place by the Public Health Service (PHS) regarding allegations of research misconduct [55]. These regulations only apply to projects funded by a PHS organization. Universities who receive funding from a PHS organization must have internal policies that reflect the requirements of these regulations. In many cases, the internal policies expand to all research and are not limited to PHS funded projects.

The ORI is the agency that enforces research misconduct rules. It provides a sample policy that gives great detail and guidance when creating a research misconduct policy. Following is a detail of the requirements the internal policies should contain.

7.2 Research Misconduct

When an accusation of research misconduct arises, it is normally a unique case. There is no way to predict exactly what the issue may be. The parties and the university itself may be very affected by an accusation. A finding of research misconduct can lead to suspension of one's career as well as distress of reputation. It is important that all of these cases are treated carefully and follow the orders of procedure outlined by any policy [56].

Many issues may arise regarding research, but not all of them will constitute research misconduct. The ORI defines research misconduct as fabrication, falsification, and plagiarism. Fabrication is making up data or result and recording or reporting them. Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record. Plagiarism is the appropriation of another person's ideals, processes, results, or works without giving appropriate credit [57]. The key to defining research misconduct is intent. Research misconduct does not include honest error or differences of opinion. Research misconduct must consist of intentional fabrication, falsification, or plagiarism.

7.3 Parties

When following the guidelines of the ORI, it is easiest to become familiar with its terminology and incorporate the mechanism into internal policies. These are the generally accepted terms used across PHS-funded institutions.

7.3.1 Research Integrity Officer

The Research Integrity Officer is often referred to as the RIO. The RIO is an institutional official who is responsible to the initial assessment of research misconduct allegations [58]. His or her assessment will determine whether the allegations warrants further review. If an accusation moves forward, the RIO will oversee the entire investigation and inquiry. A committee will be formed by the RIO, who will take the lead in assessing the information. The RIO will collect and securely store all evidence gathered and serve as the key point throughout the entire proceeding. Institutions vary in regards to who serves the role of RIO. The individual is often a member of the Division of Research or similar office. In smaller institutions, the RIO may also be a faculty member, but this is not always the case. The RIO's identity should not be kept secret. It is important that the entire university, including undergraduates, is easily able to ascertain who the RIO is.

7.3.1.1 RESEARCH INTEGRITY BOOT CAMP

The ORI offers a boot camp for RIOs. Many different aspects of the RIO position are covered in great detail. It is extremely helpful for new RIOs. The boot camps are not regularly scheduled. If someone is interested in attending they should contact the ORI for further details.

7.3.1.2 ASSOCIATION OF RESEARCH INTEGRITY OFFICERS

In 2013 the first Association of Research Integrity Officers meeting was held. This is designed to be an organization separate from ORI, where RIOs can stay up-to-date on issues affecting research misconduct. The organization is in early stages of development, but as membership grows it is expected to have a large impact on the RIO community and strengthen the resources available to RIOs.

7.3.2 Deciding Official

The deciding official (DO) is an institutional official who is responsible for final determinations of research misconduct [58]. Generally, they are removed from the inquiry and

investigation phases. They are informed of the findings of the investigation and will make the final call based on the information provided. In certain institutions, they are responsible for listening to any appeals regarding research misconduct findings. The DO will decide what administrative actions should be taken and ensuring they are enforced. This often may be detailed as other university policies and procedures will come into play, such as any union agreements, tenure policies, and human resource requirements. The DO and the RIO must never be the same person; however, in many cases, the DO has the right to appoint individuals to serve on a committee.

7.3.3 Complainant

Researchers do not draw attention to their own instances of misconduct. The complainant, sometimes referred to as the whistleblower, is the person who notifies the institution of the potential misconduct [59]. Organizations should encourage complainants to come forward and make the process as easy as possible. Some organizations allow for anonymous complainants. If this is the case, then the RIO will act as the complainant for purposes of record.

The first person the complainant should speak with is the RIO. Often potential complainants will go to a dean or department chair. These individuals should immediately refer the complainant to the RIO instead of trying to handle this issue themselves. It is in the best interests of the complainant to study any internal policies before making an accusation. The policy should let them know who the RIO is, where the allegation should be reported, what they can expect from the process, and what protections they are provided. After misconduct proceedings start, the complainant does not carry any additional responsibilities. They may be called to investigations in the role of witness, but they will not be considered a party to the proceedings [58].

7.3.3.1 RETALIATION

In accordance with 42 CFR 93, organizations must "take all reasonable and practical steps to protect the positions and reputations of good faith complainants, witnesses and committee members and protect them from retaliation by respondents and other institutional members."

It is important that the allegation is made in good faith for this provision to come into play. The complainant must have an honest belief that there is an instance of research misconduct. When there is an allegation made with reckless disregard or willful ignorance of facts, these protections are not awarded to the complainant.

If the accusation is in good faith it is the responsibility of the university to ensure that there is no retaliation taken against the complainant (ORI, http://ori.hhs.gov/guidelines-whistleblowers [accessed 12.17.14]).

7.3.4 Respondent

The Respondent is the party accused of research misconduct [60]. All allegations should be taken seriously as even an allegation can damage a researcher's reputation and career. The majority of allegations of research misconduct will not stop after initial review by the RIO, but once the respondent is informed of the accusation, he or she should start by reviewing all policies, both internal and external.

It is recommended that the respondent start to identify witnesses, gather documentation, decide whether or not an expert should be consulted, make sure that no retaliation is taken against whoever he or she believes the complainant to be, and try to maintain confidentiality of the proceedings.

The respondent should receive a fair and objective investigation. They should also be given an opportunity to admit to research misconduct and avoid the procedure. They must be given an opportunity to comment on the findings, a copy of any reports, access to the evidence, and be notified throughout the stages of the proceedings (42 CFR 93).

If there is no finding of research misconduct the University should take all steps possible to restore the reputation of the respondent (ORI, Survey of Accused but Exonerated Individuals in Research Misconduct Cases [accessed 12.17.14]). If the respondent leaves the institution before the proceedings are over, the proceedings will continue. Resigning or transferring will not stop any research misconduct cases.

7.4 Initial Allegations

The institution's policy should outline methods that initial allegation may be made [61]. This should include whether anonymous allegations are accepted and what form the allegations must take, whether it be oral or written. Individuals may make complaints directly to ORI or the PHS funding institutions.

7.5 Initial Assessment

The initial assessment phase is where most accusation of research misconduct cease [58]. The RIO will meet with the complainant, whether face-to-face or by electronic means. After hearing the allegation, the RIO may conduct interviews, analyze documents, and any other investigation activities deemed necessary.

The RIO will decide whether or not the allegation falls within the scope of research misconduct. He or she will also determine how credible the accusation is and whether or not there is potential evidence. If the RIO believes the allegation is credible and fits the definition of research misconduct, the case will move forward to the inquiry phase.

7.6 Inquiry Phase

The inquiry phase is not designed to determine guilt, rather to make an evaluation of the available evidence, including witnesses, and determining whether there is evidence to move forward to the investigation phase [58].

7.6.1 Inquiry Committee

At this point, the RIO will establish the inquiry committee. The internal university policy should reflect how many members serve on the committee, usually ranging from three to six. They may be institutional officials, experts in the field under inquiry, or faculty members. It is normally at the discretion of the RIO who forms the inquiry committee.

7.6.1.1 SEQUESTRATION

The respondent must be notified that the accusation is moving toward an inquiry. It is normally in the RIO's best interests that any sequestration of computers, notebooks, and any other possible evidence be sequestered at this time. Evidence that is valuable to the inquiry and investigation should be sequestered as efficiently and speedily as possible.

It is important that procedures are in place for securing the original documentation and any other records that are relating to the research. Items can be lost or destroyed that will greatly hinder the inquiry and further investigation phases. This process may be intrusive and stressful for the respondent and the university as well.

ORI is available to offer any technical assistance and a variety of forensic tools that may be used to evaluate the evidence [62].

7.6.2 Final Report

After the inquiry phase, a report should be generated detailing the findings. This should summarize the evidence, list the witnesses, and indicate whether or not an investigation is warranted. If an investigation is warranted, then justification should be provided.

PHS regulations mandate that the inquiry report be finished within 60 days from the commencement of the inquiry. The inquiry report must be submitted to ORI [63].

7.7 Investigation Phase

If the inquiry committee recommends the proceedings move forward, the investigation phase will take place next. The investigation will explore the details and examine all the evidence in depth. It is necessary that the investigation broaden the scope of the inquiry committee and determine whether there are any additional concerns regarding research misconduct [58].

The investigation will consist of a variety of activities. This can include, but are not limited to, reviewing the evidence, reviewing published materials, inspecting laboratories, interviewing parties, and pursuing all leads regarding the conduct in question.

7.7.1 Investigation Committee

In a similar manner to the inquiry phase, an investigation committee will be brought forward. It is at the discretion of the university policy whether the inquiry committee and investigation committee have the same makeup.

7.7.2 Investigation Report

The report must include the nature of the allegation, the documentation of all PHS support (if applicable), the specific allegation of research misconduct, the institutional policies and procedures under which the proceedings were conducted, a summary of the evidence, a list of the evidence taken into custody, and a statement of the findings for each allegation.

The DO will review the investigation report and has the final decision on whether or not to accept the findings. If the research is PHS funded, the investigation report must be submitted to ORI for review as well as the PHS funding agency [58].

If the research is PHS-funded and if the investigation cannot be completed in 120 calendar days, then the RIO will submit a written request to the ORI for an extension. If the request is granted, the university will file periodic progress reports as requested by the ORI.

7.8 Institutional Decision

Once the decision has been made on whether or not research misconduct occurred, the institution must inform the respondent and the complainant of the final determination in writing [57].

The DO must also determine what other organizations must be notified. This could include publications, licensing boards, professional societies, or law enforcement.

7.9 ORI Oversight

Both the inquiry and investigation reports must be sent to ORI for final review. ORI will examine both documents carefully and request any subsequent data or information necessary to complete their overview.

Once the overview is completed, ORI will draft an oversight report. If the allegation of research misconduct was supported, then ORI will try to work with the respondent to create a Voluntary Exclusion Agreement in which the respondent accepts impositions based on the misconduct. If an agreement cannot be reached, ORI will recommend impositions of administrative actions to the Assistant Secretary for Health or submit a charge letter to the HHS Departmental Appeals Board.

7.10 Appeals

7.10.1 Institutional Level

It is up to the institution whether or not an internal appeal process will be created. There is no requirement for an appeal to exist from the ORI [58]. If an appeal option exists, the institution must include this information in its policy. This includes the grounds and procedures for filing an appeal and who may file the appeal.

7.10.2 ORI

A respondent may request a hearing on the PHS finding or administrative action before the HHS Department Appeals Board within 30 days of receipt of the final report from the ORI. This is conducted by an administrative law judge.

7.11 Case Summaries

ORI offers case summaries of research misconduct proceedings on its Website. The case summaries can be very informative for RIOs and others interested in research misconduct processes [64].

References

[1] Atkins Daniel, Droegemeier Kelvin, Feldman Stuart, Garcia-Molina Hector, Klein Michael, Messerschmitt David, et al. Revolutionizing science and engineering through cyberinfrastructure: report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure. National Science Foundation; 2003.

[2] H.R, 708 – 113th Congress (2013-2014). http://beta.congress.gov/bill/113th-congress/house-bill/708 [accessed 02.02.14].

[3] NOT-OD-03-032, National Institute of Health. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html [accessed 12.12.13].

[4] PL 110–161, Section 218. http://publicaccess.nih.gov/policy.htm [accessed 12.12.13].

[5] Interagency Working Group on Digital Data. Harnessing the power of digital data for science and society. Executive summary. Committee on Science of the National Science and Technology Council; 2009.

[6] Chapter II.C.2.j. Grant proposal guide. National Science Foundation; 2011.
[7] http://www.whitehouse.gov/administration/eop/ostp/initiatives#Openness [accessed 27.11.13].
[8] Holdren John P. Section 4b Increasing access to the results of federally funded scientific research. Office of Science and Technology Policy; 2013.
[9] Clinical Tools Inc. Guidelines for responsible data management in scientific research. 2006.
[10] Kalichman Michael. Data management. In: Kulakowski EC, Chronoister LU, editors. Research administration and management. 2006. Sudbury, Massachusetts.
[11] Stenech Nicholas H. Introduction to the responsible conduct of research. Office of Research Integrity; 2000.
[12] http://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dotopic.html [accessed 02.12.13].
[13] Dietrich Dianne, Adamus Trisha, Miner Alison, Steinhart Gail. De-mystifying the data management requirements of research funders. ISTL Summer 2012. http://www.istl.org/12-summer/refereed1.html. [accessed 17.12.14].
[14] Chapter VI.D.4. Awards and administration guide (AAG). National Science Foundation; 2013.
[15] Updated information about the data management plan required for all proposals. http://www.nsf.gov/bio/pubs/BIODMP061511.pdf [accessed 14.12.13].
[16] Data management for NSF engineering directorate proposals and awards. http://nsf.gov/eng/general/ENG_DMP_Policy.pdf [accessed 14.12.13].
[17] Data management for NSF EHR directorate. National Science Foundation; March 2011.
[18] Data management guidance for CISE proposals and awards. http://www.nsf.gov/cise/cise_dmp.jsp [accessed 14.12.13].
[19] Directorate for Geosciences—Data policies. http://www.nsf.gov/geo/geo-data-policies/index.jsp [accessed 14.12.13].
[20] Data management for NSF SBE directorate proposals and awards. October 2010.
[21] Dissemination and sharing of research results. http://www.nsf.gov/bfa/dias/policy/dmp.jsp [accessed 14.12.13].
[22] Directorate of Mathematical and Physical Sciences Division of Physics Advice to PIs on Data Management Plans. http://www.nsf.gov/bfa/dias/policy/dmpdocs/phy.pdf [accessed 14.12.13].
[23] Directorate of Mathematical and Physical Sciences Division of Mathematical Sciences Advice to PIs on Data Management Plans. http://www.nsf.gov/bfa/dias/policy/dmpdocs/dms.pdf [accessed 14.12.13].
[24] Directorate of Mathematical and Physical Sciences Division of Materials Research Advice to PIs on Data Management Plans. http://www.nsf.gov/bfa/dias/policy/dmpdocs/dmr.pdf [accessed 14.12.13].
[25] Directorate of Mathematical and Physical Sciences Division of Chemistry Advice to PIs on Data Management Plans. http://www.nsf.gov/bfa/dias/policy/dmpdocs/che.pdf [accessed 14.12.13].
[26] Directorate of Mathematical and Physical Sciences Division of Astronomical Sciences Advice to PIs on Data Management Plans. http://www.nsf.gov/bfa/dias/policy/dmpdocs/ast.pdf [accessed 14.12.13].
[27] PL 110-161, Div G, Title II, Section 218. http://publicaccess.nih.gov/policy.htm.
[28] NOT-OD-03-032. Final NIH statement of sharing research data. February 2003.
[29] Office of the Chief Information Officer. http://energy.gov/cio/office-chief-information-officer [accessed 14.12.13].
[30] Biven Laura. Office of Science Statement on Digital Data Management. United States Department of Energy; March 2013.
[31] Climate and Environmental Sciences Division. http://science.energy.gov/ber/research/cesd/data-management/ [accessed 17.12.13].
[32] Hook, L.A., Christensen, S.W. Developing data management policy and guidance documents for your NARSTO program or project. http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf [accessed 17.12.13].
[33] Policy statement on data sharing in IES research centers. http://ies.ed.gov/funding/datasharing_policy.asp [accessed 17.12.13].
[34] U.S. Environmental Protection Agency. Open government data quality plan 1.0. May 2010.
[35] U.S. Environmental Protection Agency, Office of Research and Development, Office of Science Information Management. Survey of EPA and other federal agency scientific data management policies and guidance. April 2010.
[36] Meeker Sherwin. ADS chapter 203 assessing and learning. Sections 203.3.1.8; 203.3.2–203.3-17. November 2012.
[37] USAID/DCHA/ASHA Annual Program Statement. II 3.6. Monitoring and evaluation plan. November 2013.
[38] National Oceanographic Data Center. Long version of the data submission guidelines. http://www.nodc.noaa.gov/General/NODC-Submit/submit-guide.html#polguide [accessed 17.12.13].

[39] Guch Ingrid. It's time to share: proposed NOAA data policy. September 2011.
[40] Data rights and related issues. http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/data-rights-related-issues/ [accessed 18.12.13].
[41] Earth Science Division, NASA Science Mission Directorate. Guidelines for development of a data management plan. Version 1.0. January 2011.
[42] Fisher Richard R. NASA heliophysics science data management policy. Version 1. April 2009.
[43] Data management plans for NEH office of digital Humanities proposals and awards. http://www.neh.gov/files/grants/data_management_plans_2013.pdf [accessed 18.12.13].
[44] American Heart Association. Guide for affiliate research awards. December 2011.
[45] Alfred P. Sloan Foundation. Digital information technology. http://www.sloan.org/major-program-areas/digital-information-technology/data-and-computational-research/ [accessed 18.12.13].
[46] CDC-GA-2005-14. CDC/ATSDR policy of releasing and sharing data. September 2005.
[47] National Institute of Justice, Data archiving plans for NIJ funding applicants. http://www.nij.gov/funding/data-resources-program/applying/Pages/data-archiving-strategies.aspx [accessed 19.12.13].
[48] National Institute of standards and technology guidelines, information quality Standards, and administrative mechanism. http://www.nist.gov/director/quality_standards.cfm [accessed 19.12.13].
[49] United States Department of Agriculture, Grants. http://www.csrees.usda.gov/funding/electronic.html [accessed 19.12.13].
[50] ORI. https://ori.hhs.gov/data-management-0 [accessed 01.12.13].
[51] NOT-OD-03-032, National Institute of Health. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html [accessed 12.12.13].
[52] University of California. https://dmp.cdlib.org/ [accessed 17.12.14].
[53] Dietrich D, Adamus T, Miner A, Steinhart G. De-mystifying the data management requirements of research funders. 2012.
[54] Jahnke LM, Asher A. Dilemma of digital stewardship a: research ethics and the problems of data sharing. In: Research data management. Council on Library and Information Resources; 2013.
[55] Responsibility of PHS Awardee and Applicant Institutions for dealing With and Reporting Possible Misconduct in Science, 42 CFR 50, 93. 2005.
[56] ORI. http://ori.hhs.gov/handling-misconduct [accessed 12.01.14].
[57] Responsibility of PHS Awardee and Applicant Institutions for dealing With and Reporting Possible Misconduct in Science, 42 CFR 93. 2005.
[58] ORI, Sample policy and procedures. http://ori.hhs.gov/sites/default/files/SamplePolicyandProcedures-5-07.pdf [accessed 12.01.14].
[59] ORI. http://ori.hhs.gov/complainant [accessed 12.01.14].
[60] ORI. http://ori.hhs.gov/respondent [accessed 12.01.14].
[61] ORI. http://ori.hhs.gov/handling-allegations [accessed 12.01.14].
[62] ORI. http://ori.hhs.gov/forensic-tools [accessed 12.01.14].
[63] ORI. http://ori.hhs.gov/sites/default/files/SamplePolicyandProcedures-5-07.pdf. The inquiry report [accessed 12.01.14].
[64] ORI. http://ori.hhs.gov/case_summary [accessed 12.01.14].