Senate Research Committee
Monday, October 25, 2010
2:15 - 4:00
238A Morrill Hall

Present: Thomas Vaughan (chair pro tem), Breanna Byiers, Arlene Carney, Margaret Catambay, Jerry Cohen, Donald Dengel, Seung-Ho Joo, Frances Lawrenz, Jennifer Linde, Timothy Mulcahy, Federico Ponce de Leon, LaDora Thompson, Karen Williams

Absent: Melissa Anderson, Mustafa al'Absi, Anna Clark, Paul Cleary, Robin Dittman, Tucker LeBien, Toni Leeth, Kola Okuyemi, Lynn Zentner

Guests: Professor Peter Bitterman (Faculty Consultative Committee)

Other: Peggy Sundermeyer (Office of the Vice President for Research)

[In these minutes: (1) Responsible Conduct of Research policy; (2) data retention and management]

1.   **Responsible Conduct of Research (RCR) Policy**

        Professor Vaughan convened the meeting at 2:15, explained that Professor Anderson was out of town, and welcomed Professor Bitterman from the Faculty Consultative Committee (FCC) to discuss the recently-approved policy on RCR.

        Professor Bitterman noted the emails that FCC members had exchanged about the policy, copies of which had been provided to the Committee. He explained that he has been at the University since 1985 and has had most kinds of grants that one can have. His contribution to the discussion today is to highlight the challenges to being prudent investigators in following the RCR policy—a policy that he strongly supports. His point is that administrative technicalities in the policy make things difficult for PIs. Issues arise with respect to IRB/IACUC procedures, training modules, co-investigators, employees, and with outside consortia.

        Even if one is up to date in training, Professor Bitterman said, one receives a burdensome string of emails when applying for a grant, many of them unrelated to what he does and wasteful of the University's time. As the vice chair of a large research department, as he is trying to meet grant deadlines of his own, he must also deal with junior faculty terrified about deadlines. There is, he said, a whole string of administrative events related to grants that could be eliminated.

        Moreover, the dates that training expires are random. The average active PI has five; his suggestion is that PIs be able to pick a date when they need to be updated. Professor Cohen said he understood that one can change the anniversary date by taking training early; the date of that training then becomes the anniversary date. So what Professor Bitterman is asking for can be done. Where does one find the dates, Professor Bitterman asked? The current training record has anniversary dates, Professor

---

Cohen responded; one chooses a preferred date and it becomes the anniversary date. True, Professor Bitterman agreed, except that not everyone needs all of them. PIs can fix the dates themselves, but from an institutional point of view, it would be a lot more efficient to make the process regular. This is burdensome because people should focus on RCR issues. Does the selection of a date also include the discussion component also, Professor Linde asked? The PI can't regulate when the relevant discussion sessions will be offered.

Professor Bitterman agreed. He said he would like to see the process more regularized so it is easier and more assistance with tracking so it is also easier. The whole point of his presence, he said, is to suggest the Committee think about what be regularized and made easy for a new person to follow the procedures.

Separate approvals for every grant make things more difficult as well, Professor Bitterman said. If one has different start and stop times, and different stop times with IRB and IACUC approval, it is a problem to keep track of them. If one has only one grant, it is not a big problem, but for anyone with two or more grants, it is, even if the procedures are the same for both grants. There is a risk for the University, because sometimes people believe they are in compliance when they are not, and there is a waste of time for people and the University when sometimes they believe they are not in compliance but they are. Even a prudent person can get in trouble, given the way things are tracked.

Professor Vaughan said he has several grants and has stand-alone protocols with the IRB and IACUC, and he needs to update them from time to time.

Professor Thompson said this is a complex issue and extends into the regulation of research laboratories and shared research areas. There are multiple regulatory units on campus that are inspecting the same research areas. Hence, PIs are being inspected multiple times a year and there is redundancy in the inspections of shared research spaces. A database may increase the transfer of inspection results decreasing the redundancy of inspections.

Vice President Mulcahy said he can understand the frustration, although he does not believe it is as difficult as some believe. They are trying to resolve the problems and trying to implement a research compliance database, something on which the University has not done a great job; they hope to have it up and running within the year, with one site one can log into. What they need are specific recommendations on how to do this; it would help if the Committee came up with specific recommendations.

In terms of inspections, Vice President Mulcahy agreed and said he has a group working on trying to cut the number and get the inspections done all at once. The problem is different domains with different responsibilities. And some PIs will not want five people all at once over two days; they would prefer the inspections be spread out. He also agreed with Professor Bitterman that the data should all be in one place. Professor Bitterman surmised that if one were to poll PIs, most would prefer to pick a date for inspections rather than face recurring disruptions.

Professor Bitterman suggested there be a "certified PI" program so one is done when trained. It should indicate the limits and scope of responsibilities: Consortium? Employees? The PI could do it on his or her own schedule. By the time a grant is received, Vice President Mulcahy observed, the federal government expects the training to be done. If it is not complete, there is a problem. Professor Bitterman agreed and said one should receive an email reminding him/her to get it done. There is also a nesting

problem; is one responsible for a co-PI? A single-site relational database and certified PI program that would allow one to see if a co-PI had completed training would be a great step forward, he said. Professor Vaughan said it would be helpful to have data in one source, a place one could enter by name and see everything one is responsible for and the funding one has.

That was his vision when he came to the University, Vice President Mulcahy related, but so many units own different pieces of data that the vision has been difficult to accomplish. He said he cannot vouch for anything financial because those data fall outside his realm of control, but compliance and information on grant cycles and protocols they should be able to assemble.

Professor Cohen reported that he serves on the Federal Demonstration Partnership and has learned that these problems are universal He suggested that with respect to grants.gov, the University might find example of places that are doing well—although he has the sense that there are not too many good examples; some things are driven by regulations. It may be true that there is no list of institutions that follow best practices, Vice President Mulcahy said, but if there were, Minnesota wouldn't be on it. The University has a very cumbersome process. He said that Professor Bitterman's suggestion of a "certified PI" is an interesting one, with the status indicating the person is caught up on all training. They would need to be careful, because some people want "just in time" training, but they are open to exploring ways to adopt the suggestion. He said he hoped that the new systems would also eliminate some of the problems.

One problem that could easily be corrected, Professor Cohen said, is that many emails are not clear and are more ominous than they need to be. They could have a header, indicating one is receiving them because one has submitted a grant application and it is a good idea for the person to read it. Vice President Mulcahy agreed that the messages could be toned down from the "hand of God" tenor they now carry; they could be more civil and supportive.

Another point, Professor Cohen said, is that IRB and IACUC deal with important issues, but he fears they have become too institutionalized, so there is not enough discussion and thinking about the issues. The approach they take is too focused on regulation and they do not find any indication of knowledge. Professor Bitterman agreed; if something is on the list, it is something to check off, but these are very important matters. The Medical School faculty feel they are important and they want to do things right, not just check off a box. This is a big deal because these people are their patients. Generating 47 emails, however, achieves the result opposite from what the University intends. The idea is that one should be interested because it is important to one's work and career, Vice President Mulcahy agreed, but he observed they can encourage but they cannot force, and then the faculty blame them for not telling they had to do something. He said the University needs a system that targets plus or minus two SDs and not the ends of the spectrum—and should reserve the "hand of God" message for those at the ends. To be a "certified PI" should be the goal of everyone who gets grants at the University; there are specific federal regulations that determine when one must do IRB or IACUC training, but the University should be able to certify when someone is up to date in required training. He said he would take these ideas to the faculty committee that advises the Responsible Conduct of Research committee.

Professor Cohen said that there are certain things he must get approval for that require going through a lot of forms that do not apply to him. He never works with animals, for example. They could do a lot to target the minimum experience required or a protocol on file. Dr. Mulcahy said he would be

happy to forward ideas to IACUC.   It has completely changed its operation; they would be interested in the suggestions from the Committee.

One difficulty, Dr. Mulcahy told the Committee, is getting knowledgeable faculty to serve on committees, people who can identify what is ridiculous and what should be changed. If one has the opportunity to serve, he asked that they please do so, because that is where faculty can have the greatest impact and where they can see significant changes. The changes in the animal-use protocols occurred because three active faculty members got on the committee and brought PI views.

What's the distillation of his views, Professor Vaughan asked Professor Bitterman? A certified PI program and a database with all the information a PI needs in one place, Professor Bitterman replied. Vice President Mulcahy suggested the Committee might adopt a motion asking his office to look into the possibility of a "certified PI" program and come back with a report. The Committee voted unanimously in favor of the motion and agreed that Dr. Mulcahy would report back in about two months.

Professor Vaughan thanked Professor Bitterman for joining the meeting.

2.   Data Retention and Management

Professor Vaughan now welcomed University Librarian Wendy Lougee and Vice President Stephen Cawley to the meeting to discussion the new NSF requirement for a data-management plan that takes effect in January, 2011.

Ms. Lougee distributed copies of a handout and highlighted the main points. The handout was this:

Data Management Plan
NSF Data Management Plan Web site: http://www.nsf.gov/bfa/dias/policy/dmp.jsp

1. What is it?

Beginning January 18, 2011, the National Science Foundation will require a Data Management Plan (DMP) included with all *new* proposals. (This does not pertain to suppl. support to an existing award). This supplementary document, of no more than two pages, will describe how the proposal will conform to the NSF data sharing policy. The updated Grant Proposal Guide (GPG) Chapter II.C.2.jsuggests that a DMP include:
        1. the types of data (including samples, physical collections, software);
        2. metadata standards to be used;
        3. policies for access and sharing (including provisions for privacy/intellectual property);
        4. policies and provisions for re-use; and
        5. plans for archiving and preservation of access.

Important points:

--   Data management requirements and plans specific to NSF directorates and programs should be followed (e.g., Engineering Directorate (ENG) DMP Policy (pdf)). If guidance specific to the program is not available, then the requirements above apply.

--   Fastlane will not permit submission of a proposal that is missing a Data Management Plan.
--   A valid Data Management Plan may include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification.
--   The Data Management Plan will be reviewed as an integral part of the proposal, coming under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance.

2. How is it being implemented?

NSF has long had a sharing of data requirement that expects investigators to share with other researchers, "at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work." The *updated* requirement (Jan 2011) goes on to say, "Grantees are expected to encourage and facilitate such sharing" and that NSF will implement these policies for dissemination and sharing of research results through:

1. the proposal review process;
2. award negotiations and conditions; and
3. appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like.

Discipline Guidelines

Several specific programs provides guidance on preparation of data management plans, If an NSF solicitation provides guidance on preparation of data management plans, such guidance must be followed.

--   **Engineering Directorate (ENG)**: minimum retention is 3 years, data accessible immediately after publication (expect with patented information), see Directorate-wide Guidance **Geological Sciences Directorate (GEO)**:
  --   Division of Earth Sciences: preservation for data supporting long-term research is required; data accessible no more than 2 years after collection
  --   Integrated Ocean Drilling Program: ensures availability of drill samples are publicly available for access 36 months after research completion.
  --   Division of Ocean Sciences: data must be submitted to the appropriate national data center as soon as possible, (no later than 2 years after collection; metadata inventory within 60 days after the observational period/cruise.)

--   **Social, Behavioral and Economic Sciences Directorate (SBE)** ○ Data Archiving Policy for the Division of Social and Economic Sciences (SES): data, in fully cleaned and documented form, must be placed in a data archive or library within one year after the expiration of an award. Before an award is made, investigators will be asked to specify in writing where they plan to deposit their data set(s) – e.g., ICPSR.

Data Access

In many cases, the data type and disciplinary culture will determine the most appropriate place for sharing, e.g., ICPSR for social science or survey response data or NODC for oceanographic data. However, not all research disciplines have established repositories for data. In the FAQ section of the DMP policy page, the following questions are posed:

Q: There is no public database for my type of data. What can I do to provide data access?
[A.] Contact the cognizant NSF Program Officer for assistance in this situation."

Q: Should the budget and its justification specifically address the costs of implementing the DMP?
[A.] Yes. As long as the costs are allowable in accordance with the applicable cost principles, and necessary to implement the Data Management Plan, such costs may be included (typically on Line G2) of the proposal budget, and justified in the budget justification.

Q: My institution's policy is that the data and all supporting materials from all research are owned and must remain with the institution if I leave. What is the impact on DMP?
[A.] Data maintenance and archiving by an institution is one avenue by which data preservation and access can be achieved. However, the data access plan must address the institutional strategy for providing access to relevant data and supporting materials.

3. Libraries Services

The University Libraries education/consulting services include:

-- 	New web site bringing together tools and resources for "Managing Your Data"
	https://www.lib.umn.edu/datamanagement
	-- 	A list of well known data repositories in various disciplines
	-- 	A compilation of the research computing tools on campus:
		https://www.lib.umn.edu/datamanagement/computing
-- 	Since Jan 2009, workshop on "Introduction to Data Management for Scientists and Engineers."
	Recorded version available online https://www.lib.umn.edu/datamanagement/data101
-- 	In Fall 2010, two new workshops in response to the NSF grant requirement:
	-- 	**Introduction to Data Management: Best practices for managing digital research data:** introductory workshop on the tools and tips addressing practical problems (including storage options, file management, and metadata)
	-- 	**Creating a Data Management Plan for Your Grant Application:** Overview of a data management plan with a focus on data preservation issues tools/repositories for sharing data

It is likely that each directorate will develop more discipline-specific policies, Ms. Lougee related. The libraries are documenting details as released. There are no readily available data to assess the campus impact of the policy; there was a recent President's Emerging Leaders (PEL) group that surveyed data-management practices of faculty and graduate students. The findings indicated the majority of respondents produced less than one gigabyte of data per week, so relatively small numbers. About 19% of respondents produced more than that, and a very few produced terabytes of data. Few people relied on central storage systems; only 42% used department servers and many use their own laptop or desktop—and many acknowledged that they had lost data. Few used cloud services.

The libraries are trying to provide programs that differentiate between fields, Ms. Lougee said, and will have workshops on the NSF policy. They are planning to work with the Office of the Vice President for Research on guidance and potentially templates, although it is not clear if NSF will accept the use of a template from institutions.

The PEL group recommended data audits, better understanding different disciplinary needs, education, better campus options for secure infrastructure and archiving services. The last involves the ability to monitor data over time and commit to making it useful, perhaps including migrating the data to new systems or emulation actions as needed.

Vice President Cawley reported that his office has a subcommittee of college information-technology directors working with others to understand research storage needs and to design solutions. Faculty tend not to use central options, but they believe that if there will be a lot of data, there needs to be an infrastructure that works seamlessly across environments (from a supercomputer to a laptop), and the Office of Information Technology will put resources into solutions when it receives recommendations.

They are dealing with "bit rot," Mr. Cawley commented; there have been a lot of system changes and one result is that a lot of data "went away." They understand what data must be preserved and which must be translated to new systems. The University generates huge amounts of data and it is important to understand what must be kept and what need not be. Ms. Lougee said that NSF does not assume all data must be kept; the data-management plan articulates what one will do with the data and what happens if it is lost.

A fundamental question he cannot get his arms around is how NSF defines "data," Vice President Mulcahy said. He noted that he has slides and notebooks that are not on computers; people need to know what they must archive and what lengths they must go to in order to do so. Is there a guide? Ms. Lougee said they try to answer that question in the FAQs provided, but it is not clear yet how broadly NSF will define data.

A component part of the proposal is that when one presents the data-management part of the plan, one tells reviewers what data will be available and they can say whether they believe that plan acceptable, Professor Cohen said.

The community will be at a complete loss on how to handle this, Dr. Mulcahy said, and it will be as important to help people as it will be to figure out how to store things.

Why would archives need to contain more than the publications and progress reports, Professor Vaughan asked? Because they want data available to be repurposed for other research, Ms. Lougee said, but they will recognized that not all can be. Professor Cohen said that in some fields, the plan should be to publish the papers and that should be sufficient; in some fields, however, publication is a poor substitute for data, and NSF will want the data available for other uses.

Professor Cohen reported that he has five NSF grants and the problem with faculty receiving 5 gigabytes of storage space is that his grants can generate 5 gigabytes of data in an afternoon. They have not had the support they need so they built data-management into the request. He suggested working with the Minnesota Supercomputer Institute because the amount of resources needed to comply with the NSF rule is amazing. It is not easy to store data unless one is plugging into a database that everyone is using, which most faculty members do not. He said the current infrastructure at the University is not ready to accommodate the NSF policy and if they want people to be ready, they need to be able to get into the system smoothly to help archive data.

Ms. Lougee said the Libraries can do some things to help with management plans on the low end, and can help with education, but there will be scale issues and how to leverage those with experience; must each college develop expertise? Vice President Cawley reported that the Senate Committee on Information Technologies will discuss at its next meeting a requirement that every faculty member be provided 2 terabytes of storage. The most expensive storage is that which is not used; it is cheaper to provide when it is used, "just in time" storage. Rather than choosing an arbitrary number, Professor Cohen said, faculty need to be able to apply for the capacity they need. Some need little, some need a lot, and the average is not a good number.

Vice President Mulcahy said that solutions at the college level are not satisfactory, because they would all be responsible for security and for updating servers, so this is a case where central capacity makes sense. They are working with University Libraries, the Office of Information Technology, and the Minnesota Supercomputer Institute to identify solutions. He said that Ms. Lougee and her team are helping to keep them aware of what they must do, Mr. Cawley puts in the hardware, and he needs to work with the PIs. He also said the University cannot do this after the fact or do it at the 11th hour.

Professor Cohen said it would be helpful to have a contact person for PIs who are contemplating unusual data sets; helping them write up the data-management plan could help ensure that they receive the grant. Ms. Lougee said that there will be people in Dr. Mulcahy's office. Dr. Mulcahy said it would also be helpful to have FAQs that address the kinds of data and the resources available. Many people will not be computer literate about this and will not know the specifics of what they need, Professor Cohen commented; it will be important to have an infrastructure that is friendly to those without a lot of experience.

Mr. Cawley reported that they are seeing more contract requirements calling for intense security as well as very specific requirements by agency and kind of data. They would like to help people deal with the requirements, especially if they can have more than 2-3 days notice.

Professor Vaughan thanked Mr. Cawley and Ms. Lougee for their report and adjourned the meeting at 3:45.

-- Gary Engstrand

University of Minnesota