

Faculty Senate Research Committee Final Report FY13, 1 April 2013

The Faculty Senate Research Committee (FSRC) has submitted reports on Standing Charge 1 (03/05/13) and Specific Charges 1 (03/05/13) and 4 (11/15/12).

The following document includes reports on Standing Charges 2, 3, 4, and 5 and Specific Charges 2 and 3.

Standing Charges 2 and 3

Standing Charge 2

Monitor the execution of the University's Restricted Research Policy (Faculty Senate Rules and Regulations, Article IX) in handling requests for exceptions.

Standing Charge 3

Serve as the body to hear faculty appeals of research rejected by the restricted research committee or by the Vice Provost for Research, as specified in the Restricted Research Policy.

As chair of the FSRC, Celka Straughn served as an ex-officio member of the Restricted Research Committee. There were no appeals of decisions made by the Restricted Research Committee. New to policy from FY12 is that students involved need to know if there is a research embargo.

Straughn met with Doug Huffman, Chair of the Restricted Research Committee and he also presented to the FSRC to discuss the role of the RRC and in relation to the FSRC. An appeal would come to the FSRC if a faculty member disagrees with the RRC. During this meeting the FSRC suggested that perhaps it would be useful to inform faculty further about restricted research issues through some sort of training since federal laws are changing regarding restrictions and with greater severity for noncompliance (similar for using human subjects). RGS is exploring ways to create a culture of awareness of these issues. RGS will work with RRC and they will bring a plan to FSRC for review in FY14.

Standing Charge 4

Monitor the implementation of policies and procedures for determining which proposals will go forward in cases where the number of grant applications that may be submitted from the University is limited. Identify problems or concerns, and report issues and recommendations to FacEx.

There were no cases presented to the FSRC. This is an issue that comes up rarely and would be generated by RGS, most likely in the case of policy changes/review. Further information is located in the policy library. It is the FSRC's role to ensure a fair review process and that those selected have enough time to develop proposal (usually a month).

Standing Charge 5

Continue working with the ACEC (Academic Computing and Electronic Communications) Committee to process needs for and issues with computing and telecommunications for research, including recommendations for sustainable policies and procedures and monitoring ongoing and new developments in IT. Report issues and recommendations to FacEx as needed.

The FSRC had no direct communication with ACEC. Committee member Larry Hoyle volunteered to serve as the FSRC representative to ACEC and he was able to attend a fall meeting. At that meeting it was shared that the Research File Storage (RFS) had become available. This is storage that is scalable both in terms of storage volume and access speed. Research projects will be able to get up to 250GB of storage at no charge. Larger allocations are available at \$1.00 per GB per year if the data are backed up and \$0.70 per GB per year if the data are not backed up, e.g. if the storage is used for a second copy of data stored elsewhere. Central IT is also looking at ultra-low cost storage (perhaps in the \$0.08 per GB per year range).

One issue that was raised in the ACEC meeting was that of student access to some of these (and other) services. Graduate students in funded positions have easily managed access to KU Anywhere and firewall rules, but other graduate students as well as undergraduates must be manually added through a cumbersome administrative process. There are cases, though, where students participate in research without being funded by an external project. Perhaps more flexible mechanisms for a role-based authentication and authorization could be developed. It might also be useful to include issues of research data in graduate student training.

An issue for research is what is still not in place at KU: a central facility for archiving data. The RFS makes no provision for curation of the data, nor does it have any capacity to use metadata for location and reuse of the data. Another issue is that of funding for archiving data once funding for projects have ended. Funding agencies (like NSF) may have policies that data must be made available for some period after a project has ended, but they have no mechanism for providing funding for storage and curation. Some universities (e.g. University of North Carolina) have addressed these issues at an institutional level.

In response to these issues, the FSRC proposes some recommendations drafted by Hoyle. Please see Appendix A, "A Whole Lifecycle Approach to Research Data." These have been shared with ACEC, but the FSRC did not receive any feedback. The FSRC suggests that KU seriously consider these issues, particularly in light of the recent White House memorandum, "Increasing Access to the Results of Federally Funded Scientific Research" (http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf). Likely KU will be required to provide open access to data (including metadata) for future grants, therefore, it would be good to address these issues in a timely manner and gain an advantage.

Once researchers start citing data, data metrics could become part of P&T files. Furthermore, creating persistent identifiers for a data set prior to publication will be useful.

Specific Charges 2 & 3: PRO/Digital Measures and Academic Analytics (AA)

Specific Charge 2

Continue working with relevant offices and committees to ensure that the new PRO/Digital Measures system provides the customization necessary to use the system for GRF reporting.

Specific Charge 3

Continue to consider issues related to the use of Academic Analytics for external evaluations of the University and internal research evaluations for departments, programs, and faculty; the applicability of AA metrics to diverse disciplines; and the flexibility or possibility to incorporate other qualitative and quantitative metrics (such as the PRO/Digital Measures system) into AA method of analysis.

KU is still in the process of identifying what is missing from AA. The largest deficit in AA appears in the arts and humanities. AA also does not capture data that faculty have created and published. It was reported that during the fall a committee was put in place to look specifically at metrics for the arts and humanities as part of an effort to identify ways for metrics to be more relevant for all disciplines (current metrics are more applicable to math and sciences). It is important to try and develop parity across different forms of research outputs, such as classic published work, online publications and creative work, and perhaps KU could be seen as a leader in this process.

AA is seen as a useful tool for understanding research outputs at the departmental level; it can also be employed at the faculty level, but KU is not subscribing to that option. The University states that it is not using AA to compare departments internally but with like departments at other universities. For these comparisons AA allows for weights based on the NRC, however, these can be modified. Weights are discipline-specific and measured by standard deviation from mean (Z-score). While AA data do not give the full story, AA comparatives are apparently better than NRC rankings, which rely on even more outdated data. The FSRC reiterates its critique from FY12 that AA lacks information about standard deviation. There also remains the question of whether foundation grants are counted in AA.

It is anticipated that AA will be helpful in making departments and the university aware of research-related opportunities. RGS is currently using AA for funding opportunities and AA is in use for decisions by Provost's office. Training remained ongoing during FY13 for deans and chairs, who will have access to AA. Data are used by administrators for strategic decision making. The FSRC cautions that there is a risk of "big brother" syndrome and expresses some concerns about who has access and how the information can be used.

KU plans to collect data beyond what currently exist in AA, for example with PRO. During FY13 PRO continued in phased implementation, with anticipated completion during FY14. There has been some difficulty capturing production in all disciplines, but the system seems to be adaptable by different units. PRO is intended to allow for individual, faculty-level understanding of research. Potential benefits for faculty with both AA and PRO include the management of research activity/production and increased visibility of faculty work. PRO will enable researchers to locate others' research interests/expertise; a goal for individual faculty is to have this sort of access, though it may not be direct. PRO will be linked with GRF funds and outcomes.

Suggested charges for FY14

1. With regard to Standing Charge 4, the FSRC could look into ways to accomplish initial reviews electronically since getting a face-to-face committee meeting takes time and these cases are often rushed.
2. It might be useful for the FSRC in FY14 to examine how metrics from AA and PRO are being determined and used, and if these are well adapted to different disciplines and types of research activity. Additionally, the FSRC could look into how these metrics are impacting faculty research, research opportunities and the understanding of faculty research profiles. The Committee recommends continuing Specific Charge 3 in FY14 to monitor what kinds of data are coming from AA and PRO and what KU is learning.
3. The FSRC recommends that RGS and the KU Libraries look carefully into issues of research data (collection, analysis, storage, curation, etc.) and that it become a new charge for the FSRC and possibly also ACEC in FY14. (See recommendations in Appendix A.)

Appendix A

A Whole Lifecycle Approach to Research Data

There are a number of facilities on campus for the collection, analysis and storage of research data on campus, but the University has not addressed some critical issues involving the care of and responsibility for research data throughout the whole data lifecycle.

The Research File Storage (RFS) facility is a good solution for managed storage space for many ongoing research projects, with access managed through the KU Active Directory and centrally managed backup and off-campus access through KUAnywhere. Other facilities include the REDCap system for collection and storage of confidential data, a number of special purpose servers, and probably a large number of desktop or laptop computers. KU ScholarWorks, the KU institutional repository, is also increasingly used for datasets and research data that researchers wish to share and offers a persistent URL and exposure to external search engines like Google along with the option to upload descriptive metadata, code books, and publications for wide public access.

The February 22, 2013 Presidential Memorandum, "Increasing Access to the Results of Federally Funded Scientific Research" (<https://petitions.whitehouse.gov/response/increasing-public-access-results-scientific-research>), directs federal funding agencies like NSF to require data to remain accessible for some time period after the funded period. Disciplinary associations also increasingly call for data underlying publications to be made accessible. Archiving data presents a number of challenges not addressed by the availability of simple storage facilities. Storage of large datasets may generate funding challenges. Data stored in proprietary formats may require periodic conversion to newer versions. Changes in storage technology may require movement of data to new storage systems. As standards for persistent identifiers and citation of data develop, documentation of linkages between data and publications will become important.

Data require ongoing updating of a variety of types of metadata through time. Data are not usable without good information about how they were generated. Technical metadata about the structures in which data are represented are needed in order for the data to be read correctly. Other kinds of metadata are needed in order for data to be discoverable or citable. Specialized metadata are needed to ensure archival management and preservation. A growing trend is for datasets to be searchable by concepts measured, or for subsets of observations based on values of certain measures, for example survey responses by males over 60. Harmonization of related measures between data sets can be complicated. The ongoing need for accurate information about data requires active curation of the data. It also implies a need for training of researchers on planning for long-term management.

A helpful model for thinking about the different kinds of information and services needed for managing data is the Johns Hopkins University stack model for Data management described by Sayeed Choudhury:

Data Management Layers

Layers	Characteristics	Sample Actions	Researcher Implications	NSF Implications
Curation	Active and ongoing management of data through its lifecycle of interest and usefulness	<ul style="list-style-type: none"> • Provide ongoing bibliographic control for data. • Link research data to publications based on the data. • Provide tools for further analysis. • Harvest metadata for the data to share with external search engines. 	<ul style="list-style-type: none"> • Feature Extraction • New query capabilities • Cross-disciplinary accessibility 	<ul style="list-style-type: none"> • Offers competitive advantage • New opportunities for data use
Preservation	Ensures that archived data can be fully used and interpreted over time	<ul style="list-style-type: none"> • Add information to maintain the <i>viability</i>, <i>render-ability</i>, and <i>understandability</i> of data long term. • Monitor format obsolescence; migrate data to new digital formats as need. • Preserve tools and/or documentation for using the data. 	<ul style="list-style-type: none"> • Ability to use own data in the future (e.g. 5 years) • Data sharing with others 	<ul style="list-style-type: none"> • Satisfies NSF needs across directorates
Archiving	Data protection is applied to stored data, including fixity checking, and assignment of data identifiers	<ul style="list-style-type: none"> • Check for viruses as data is deposited • Establish checksum snapshots over time to ensure data has not change. • Assign a persistent identifier such as a DOI or handle. • Link metadata to data 	<ul style="list-style-type: none"> • Data is better protected. • Provides persistent identifiers for locating, sharing, and referencing data. 	<ul style="list-style-type: none"> • Could satisfy most NSF requirements
Storage	Bits on disk, tape, cloud etc. Backup and restore.	<ul style="list-style-type: none"> • Place data in networked storage • Invoke backups. • Set access protocols. 	Responsibilities for <ul style="list-style-type: none"> • Restore • Sharing • Staffing 	<ul style="list-style-type: none"> • Could be enough for now, but not near-term future

Adapted by Deborah Ludwig from the "stack model" of Data Management Layers under development by John Hopkins University and based on the definition of data curation advanced by the University of Illinois Graduate School of Library and Information Science.
 See: <http://www.clir.org/initiatives-partnerships/data-curation>

No current solution available at KU offers a complete curation system. Metadata options, for example, managed by KU ScholarWorks are currently limited to descriptive elements at the object level. Depositors may or may not include enough metadata to ensure that data can be interpreted without communication with its creator (Preservation level information in the model above).

Underlying all of these issues is the question of who is ultimately responsible for providing all of the necessary services to adequately preserve data beyond the active period of a research project. The Presidential Memorandum calls for "leveraging existing archives and fostering public-private partnerships with scientific journals relevant to the agency's research". The issue of open access for data is likely to parallel that of open access movement for publications.

Some universities have begun to consider this question. See for example the report "*Research Data Stewardship at UNC*"

http://sil.unc.edu/sites/default/files/general/research/UNC_Research_Data_Stewardship_Report.pdf, in which the University of North Carolina adopts three principles

- ***"Principle 1. UNC considers the data that are material for and resulting from academic research as a public good; preserving these data and making them available to the public within regulatory and legal constraints is part of UNC's function as a public entity.***
- ***Principle 2. Researchers/data creators are responsible for specifying data life cycle plans that comply with constraints defined by pertinent law, funding agencies, and research community practices.***
- ***Principle 3. UNC is institutionally obligated to support creation, maintenance, and execution of data life cycle plans by affiliated researchers. "***
-

The University of Kansas should address these issues and evaluate which services may be needed to support the management of research data throughout its full lifecycle.

Prepared by Larry Hoyle for FSRC.