

ICPSR Council Minutes

March 22-23, 2012

Attendees

Council Members: Rogelio Saenz (Chair), Christopher H. Achen, Marilyn Andrews, Tony N. Brown, G. Sayeed Choudhury, Susan L. Cutter, Catherine A. Fitch, John Fox, Gregory N. Price, and Ann Wolpert (Past Chair)

ICPSR Staff: JD Alford, George Alter, Rita Bantom, Bryan Beecher, Dieter Burrell, Linda Detterman, John Garcia, Peter Granda, William Jacoby, Mary McEniry, James McNally, John Marcotte, Kaye Marz, Dan Meisler, Asmat Noori, JoAnne McFarland O'Rourke, Michelle Overholser, Amy Pienta, David Thomas, Mary Vardigan, and Diane Winter

Visitor: Dr. Panuchart Bunyakiati, Assistant President for ICT, University of the Thai Chamber of Commerce and Director, University of Chicago - UTCC Research Center

Director's Report

Minutes from the October 2011 Council meeting were reviewed, voted on, and passed without changes.

George Alter introduced and welcomed Dr. Panuchart Bunyakiati, who is Assistant President for ICT, University of the Thai Chamber of Commerce and Director, University of Chicago - UTCC Research Center.

Alter also welcomed new Council members Christopher H. Achen, Princeton University; Marilyn Andrews, University of Regina; Tony N. Brown, Vanderbilt University; Susan L. Cutter, University of South Carolina; John Fox, McMaster University; Linda J. Waite, University of Chicago; and announced the new Council Chair, Rogelio Saenz, University of Texas at San Antonio.

Alter noted that membership stood at 718, which included 27 institutions added for the fiscal year. FY12 revenues were currently at \$3,340,000 surpassing the forecast of \$3,300,000. Web site activity was strong, with over 250,000 visitors and over 400,000 visits to the site in the last eight months. Site visitors typically remain on the site for over five minutes and average approximately eight page views. Most of the visits to the site were from the United States, with Canada second, and China third. Among the referring sites the second highest is the Renmin University in China. ICPSR signed an agreement with Renmin about a year and half ago to launch a new membership federation in China.

Alter gave Council members a brief review of the proposed 2013 budget. The total budget projection is \$18.5 million; however, this is calculated the way the University of Michigan calculates the budget, which includes some double counting of the recharges. Alter explained the

recharge process at the University of Michigan and ICPSR. The *real* budget is closer to \$15 million. ICPSR is currently showing a deficit for the next fiscal year of about \$.5 million, but that projection is a very conservative estimate. It assumes that none of the grant funding applied for will be awarded. ICPSR's budgets tend to start out in deficit but become stronger over the course of the year. Alter explained the funding of the "Measures of Effective Teaching (MET) Longitudinal Database" MET project. George also reviewed the proposals that were submitted and awarded since October 2011.

Alter announced some staff transitions. Nancy McGovern is now Head of Curation and Preservation Services at MIT. Robbin Gonzalez is leaving to pursue research on the legal representation of children, and Michael Shove has been appointed Manager of the General Archive. We have opened searches for an Associate Director of ICPSR, a Director of Curation Services, and a Director of Education and the Childcare Archive.

The General Archive has released some new data of interest, including: Kenya Democratization Survey Project, 2006; RAND Center for Population Health and Health Disparities (CPHHD) Data Core Series, 2000; How Couples Meet and Stay Together (HCMST), 2009-2011; and the Convention Delegates Study, 2000. ICPSR's Acquisitions area has received seven new studies from the LEADs database (a list of NSF and NIH grants that generated social science data) and has received 21 commitments to deposit data. Acquisition has also spearheaded a redesign of the ICPSR online deposit form.

The Summer Program will be larger than last year. The Summer Program will offer 78 courses (as compared to 65 in 2011) with 78 instructors and 338 participants registered as of March 22. Alter highlighted three new courses for 2012: Applied Data Science: Managing Data for Re-use; Assessing and Mitigating Disclosure Risk: Essentials for Social Science; and Military Nursing Research: Fundamentals of Survey Methodology.

Students applying to ICPSR's Summer Internship Program can now apply online. The internship program is being funded this year by the NSF Research Experiences for Undergraduates (REU) grant. There were 270 applicants this year as compared to 80-90 in previous years. The Council will meet the 2012 interns at the June Council meeting at the Michigan League.

Alter reviewed the successful OR meeting, which was held October 5-7, 2011. There were 175 attendees representing 98 institutions, and 193 people watched as the sessions were live-streamed.

ICPSR is celebrating its 50th Anniversary with a symposium featuring Dr. Elinor Ostrom on June 7 followed by a banquet to wrap up the yearlong celebration.

Alter reported that ICPSR offered a retirement incentive opportunity to 16 individuals and nine people accepted the offer. Following the retirements, ICPSR leadership conducted a review of all positions with the result that some positions previously held by retirees will be replaced with about 3.235 FTE eliminated.

Plenary on Confidential Data and IRBs

Alter presented information on issues related to disclosure risk in data, placing ICPSR activities in the context of current practices and policies about confidential data.

Disclosure Risk

Alter began with a summary of risk and harm related to disclosure. We generally assume that in conducting research about human subjects the benefits (to society) outweigh the risk of harm (usually to an individual). We promise confidentiality, which is important because people may reveal information about themselves that could cause them harm if disclosed – e.g., criminal activity, antisocial activity, medical conditions.

ICPSR and other data providers need to ensure that direct identifiers (e.g., name, SSN, address) are not disseminated. Indirect identifiers – that is, characteristics like sex, race, age, and occupation that might identify a subject when combined – are also problematic in that they pose a deductive disclosure risk. The more specific the geography, the more attention must be paid to disclosure risk.

The risk is that someone might attempt to re-identify individuals. This might occur, for example, in cases of parents trying to find out if their child uses drugs, a spouse seeking hidden income, etc.

There are two scenarios related to data “intruders” – the intruder knows the respondent and is searching for him or her based on knowledge already in hand, or the intruder does not know the respondent and is searching against some comparison database. There are an increasing number of public databases against which to compare data records.

Alter went on to say that current survey designs increase the risks of disclosing subjects’ identities as we now are seeing more geographically referenced data, longitudinal data, and multi-level data.

We might think of three aspects of protecting confidential data:

- Safe data: Modifying the data to reduce the risk of re-identification
- Safe places: Physical isolation and secure technologies
- Safe people: Training and data use agreements

In terms of making data safer, we can undertake strategies such as:

- Grouping values
- Top-coding
- Aggregating geographic areas
- Swapping values
- Suppressing unique cases
- Sampling within a larger data collection
- Adding “noise”
- Replacing real data with synthetic data

With respect to safer places, ICPSR requires detailed data protection plans from prospective restricted data users to ensure that the setting and protections for the data at the site of the analysis are appropriate. ICPSR has also built a Virtual Data Enclave (VDE) that will enable researchers to access and analyze data remotely on ICPSR servers. And finally, for data that are extremely sensitive, ICPSR operates a physical data enclave at its headquarters in Ann Arbor with a non-networked environment and research outputs vetted for disclosure risk.

Secure data services at ICPSR include the Secure Data Environment (SDE), the environment in which all data processing is done. The SDE protects against accidental data leakage, uses firewalls to isolate content, and provides virtualized workstations to access content. The VDE for external users is built on this same technology.

To ensure that people work safely with data, ICPSR requires data users to provide a research plan, IRB approval, and a data protection plan. ICPSR has also established a set of rules for behavior and requires a security pledge and an institutional signature. ICPSR is working on a disclosure risk online tutorial.

The Data Use Agreement is particularly important because it spells out the guidelines that researchers must follow in the release of statistics derived from a dataset – for example, in no case should the total for a row or column of a cross-tabulation be fewer than ten. It is also stressed that violations of the agreement will be treated as a violation of policies and procedures related to scientific integrity and misconduct. There are some severe consequences possible, including sanctions suspending research grants.

ICPSR has also developed the ICPSR Data Access Request System (IDARS), which enables ICPSR staff to set up contracts for restricted data with terms of use and contract behavior preferences. It also enables end users to apply for restricted data online and permits ICPSR User Support to manage contracts and track end users.

Alter noted that there is a gradient of risk and restriction. As the risk of disclosure in data increases, so does the risk of harm and we need to make sure our protections are in keeping with these levels.

Institutional Review Boards (IRBs)

With respect to IRB supervision of ICPSR, Alter discussed the fact that ICPSR does not report to the University of Michigan IRB because ICPSR as a repository does not interact with subjects, obtain identifiable private information, obtain informed consent from subjects, or conduct research.

Informed consent is an important area. IRB advice has often been based on biomedical informed consent models that preclude data sharing, and some informed consent statements distributed by IRBs rule out data sharing. Text to avoid includes: “The data will be shared with the research team exclusively”; “The data will be shared only in aggregate form or in statistical tables”; and “Upon completion of this project, the data will be destroyed”.

NIH now supports data sharing and states that investigators should avoid developing or relying on consent processes that promise research participants that their information will not be shared with other researchers. ICPSR has posted some sample models for informed consent statements developed in consultation with the Michigan IRB.

Some IRBs have lists of public-use data that do not require review. On most campuses ICPSR's de-identified public-use data are considered exempt from IRB approval.

Alter went on to discuss the Advanced Notice of Proposed Rulemaking (ANPRM) for Revisions to the Common Rule, which would change IRB procedures. Some of the proposals are welcome – for example, replacing “exempt” with “excused,” having minimal risk studies reviewed by IRB staff, not requiring written consent for competent adults, and streamlining some procedures, such as for studies involving multiple campuses and IRBs.

Other proposals are more problematic -- for example, it is proposed that data collected for research could only be used for a purpose covered in the consent, even if the data are de-identified. Especially problematic is that all data would be under high security with HIPAA standards extended to Common Rule Research.

ICPSR developed a response in partnership with the American Educational Research Association (AERA) and also worked with the Data Preservation Alliance for the Social Sciences (Data-PASS) on responses. As a result of the many comments received from the community, the proposed rules have not yet been issued. When they are, there will be another opportunity to comment.

Budget and Policy Committee

Council: Rogelio Saenz (Chair), Marilyn Andrews, Tony N. Brown, Gregory N. Price, and Ann Wolpert
Staff: JD Alford, George Alter, Rita Bantom, and Diane Winter

I. Strategic Plan Update

The highlights of the Strategic Plan had been presented in the Director's Report so there was no further discussion by the committee.

II. Financial and Administrative Issues

A. Financial Reports

Staff reviewed the various financial reports with the committee. Some detail was provided for the benefit of the new Council members. The first draft of the 2013 budget was presented and reflected a \$556,669 deficit, of which \$331,150 was the result of a cash flow issue related to the Gates Foundation Measures of Effective Teaching (MET) project.

Revenue will be received in FY 2012 for expenditures that will not be realized until FY 2013.

The committee also reviewed the 2013 Membership budget, highlighting that most of the expected revenue increase is the result of an increase in fees for 2013.

Staff reminded the committee that the ICPSR budget methodology incorporates an ultra-conservative approach. ICPSR does not budget for sponsored projects proposed unless it has received the award notice or is very certain the project will be funded. Based on historical data, ICPSR is likely to receive two or three additional sponsored awards, improving the budget forecast to break-even or better. In June the staff will present the committee with the FY 2013 budget for approval.

The committee was also updated on the current forecast for FY 2012 and the 2011 calendar year outcome for the Summer Program.

B. Personnel Update

Staff explained the staffing/recruiting reports found in the Council binder, highlighting that there are currently 104 staff/faculty at ICPSR and nine open positions. The committee also discussed the effect of the nine retirements on the staffing. This has enabled ICPSR to obtain new skills when positions were replaced.

C. Diversity Goals and Strategies

Staff reviewed with the committee the ICPSR Diversity Strategic Plan. There was a lot of material in the briefing book to demonstrate the history of diversity initiatives at ICPSR since the October 2005 Council Meeting.

Staff asked the committee for feedback on this new plan -- specifically, does it meet Council's objectives? The committee agreed that ICPSR should focus on the first four strategic directions of the plan and the next steps are to identify goals, identify/earmark resources, and implement the plan.

The committee discussed the pipeline for candidates and ICPSR's desire to focus efforts in this direction, particularly for data processing positions and Summer Program teaching assistants.

III. Governance Issues

A. Policy on Data Access and Web Site Support After Projects Have Ended

Staff presented the policy to the committee and reported that in the last year we have not had to apply this policy.

B. Committee to Frame Principles for Evaluating Opportunities

Staff presented the principles for evaluating opportunities, which were approved by the ICPSR Director's Group. The committee approved the principles, with the caveat to incorporate: "Advance research on groups that have been historically under-represented."

C. Shanghai University of Finance and Economics

Staff provided background information on Shanghai University's relationship with ICPSR. The Shanghai University of Finance and Economics (SHUFE) has requested approval to become an ICPSR member as a federation so that they can charge their members differential rates to participate in the SHUFE program that is akin to the ICPSR Summer Program.

The committee agreed to honor SHUFE's request with the caveat that they should not misstate the benefits of becoming an ICPSR member through the SHUFE-based federation. There are already two other ICPSR federations in China (at Peking University and Renmin University), and universities have the option of joining ICPSR through those agreements.

D. ICPSR Response to ANPRM Human Subjects Research Protections, "Enhancing Protections for Research Subjects and Reducing Burden, Delay and Ambiguity for Investigators"

This information was not discussed during the Budget and Policy Committee meeting, although the committee did note that this information should be passed along to our membership institutions.

E. ICPSR Response to RFI on "Public Access to Digital Data Resulting from Federally Funded Scientific Research Office of Science and Technology Policy"

This information was not discussed during the Budget and Policy Committee meeting, although the committee did note that this information should be passed along to our membership institutions.

F. China Data Center

Staff provided the committee with background information and an update. ICPSR is currently waiting on final approval to host the China Data Center from the University of Michigan's Vice President for Research, Stephen Forrest.

IV. Projects

A. Grant / Contract Applications submitted since October 2011

This information was not discussed during the Budget and Policy committee meeting.

B. Grant / Contract Awards received since October 2011

This information was not discussed during the Budget and Policy committee meeting.

Collection Development Committee

Council: Christopher H. Achen (Chair), Susan L. Cutter, Catherine A. Fitch, and John Fox
Staff: Peter Granda and Amy Pienta
Visitor: Panuchart Bunyakiati

I. Strategic Plan Update

Peter Granda gave an update on progress in meeting the aims of the Strategic Plan.

ICPSR is a member of IFDO (International Federation of Data Organizations); ICPSR will attend the upcoming meeting of IFDO that will occur in conjunction with IASSIST.

Dr. Panuchart gave an overview of data sharing in Asia. Data sharing in Asia is achieved typically through national data centers.

Granda attended a meeting in the fall to discuss the dissemination of data from the National Academies Data Base Assessment of Research–Doctorate Programs in the US. ICPSR hopes to get these data in the future. It is unclear if the data will be public-use or restricted.

ICPSR staff attended a meeting in the fall to discuss the possibility of establishing an archive in collaboration with the Society for the Study of Nanoscience and Emerging Technologies.

ICPSR appointed an Archive Manager for the General Archive, Mike Shove. This puts the General Archive in line with the staffing structure of the topical archives.

II. Review of Current Collection Development Projects and Plans for 2012

Granda gave a report on the projects under the umbrella of Collection Development, including all of the topical archives.

A question was raised about data that are continually updated/changing. Granda explained that ICPSR typically attempts to archive data that is in final form whenever possible.

Another question dealt with how ICPSR differentiates itself from other US archiving options that are out there. ICPSR staff described that ICPSR is one of the few social science data archives in the US that offers curation of data, restricted data archiving, and other full services.

A comment was made about the difficulty of finding the 2008 election returns and turnout by county. There is a lot of historical election return data that is available but not readily accessible by the research community because it is collected and reported separately by each state. Some

researchers are particularly interested in precinct-level data, which would be even more difficult to obtain but of real value to political scientists. Obtaining these data would be a major undertaking and expense. ICPSR has traditionally invested in election return data over the years but has not had the resources lately to continue to add to its series of county-level election returns.

A question was raised about whether ICPSR ever turns away data. ICPSR staff discussed the use of the Collection Development Policy to determine what is in and out of scope for ICPSR's collection. Only a few datasets are turned away.

A suggestion was made that ICPSR look into disseminating data in an XML format that would be readily compatible with R, and there was general support for this idea. It was pointed out that ICPSR is interested in pursuing this option in the future.

III. Update of Data Management Plans (post NSF DMP requirement)

Amy Pienta gave a report on the help ICPSR is giving to NSF applicants in preparing data management plans.

The committee discussed the issue of whether ICPSR met the NSF requirement that data be available for open access, given that data with no explicit funding set aside for archiving are processed using members revenues and then made available only to members (although data can be purchased at \$500 per dataset).

The committee asked about the reach of ICPSR at the four-year college level and other less research-intensive institutions.

IV. Qualitative Mixed Method Data

Pienta gave a report on efforts made to acquire qualitative and mixed methods data.

It was suggested that ICPSR contact professional associations about datasets we should be focusing on and also make connections to collaborative data collections. Council members also suggested new datasets to acquire.

Membership and Marketing Committee

Council: Catherine A. Fitch (Chair), Marilyn Andrews, Tony N. Brown, and Rogelio Saenz
Staff: Linda Detterman, Dan Meisler, and David Thomas

Linda Detterman gave a brief overview of the committee charge and provided some background for new Council members.

I. Strategic Plan Update

Progress towards the Strategic Plan Direction 2, which focuses on membership, was reviewed.

II. Membership Activity and Utilization Reports

A. Membership Activities

Detterman reported that there were a number of institutions that were unsure about the ability to sustain their membership for the fiscal year. Southern University membership, for example, is being subsidized because of Katrina effects and the economic downturn.

B. Receivables

Detterman indicated that ICPSR had already hit its revenue forecast from membership. As of the beginning of February, ICPSR had collected approximately \$3,350,000 in member fees with about \$289,000 outstanding.

Detterman reviewed the membership and fee structure. Of particular note was the explanation of the “Italian” model (IM) of membership on which the federation system has been based since 2005. This system for international membership requires a hub member paying an initial fee of \$2,000 and any additional member paying \$500 when joining through the international federation. (Note these are the original fees – not those resulting after any increases post 2012.) This arrangement makes the payment structure for all international federations (since 2005) a standardized approach resulting in a simpler and more equitable fee structure.

About a dozen non-Italian model memberships are still in existence. These memberships negotiated an original fee for the country and that fee has increased (inflationary) over the years. Effectively, any university within those countries is welcome to affiliate with ICPSR as part of that country’s federated membership.

C. Utilization

There was a slight decrease in utilization overall. Detterman tracks institutions with low utilization numbers, follows up with them, and when possible, schedules campus visits.

Membership and utilization can be driven by any number of factors -- for example, at St. Olaf College a student was the driving force behind the membership.

III. Membership Outreach

A. 2011 OR Meeting

The group reviewed the last Official Representatives (OR) meeting, which was held in Ann Arbor in October 2011. At least 60 percent of the attendees were first-time attendees.

The meeting participants commented that virtual meetings are great but in-person meetings allow for more networking and help ORs increase awareness and pursue advocacy efforts at their home institutions. There were almost 200 hundred (193) views of the live streaming feeds of OR Meeting sessions.

B. 50th Anniversary Events

Council members were reminded to respond to the invitation for the 50th anniversary banquet, which is being held June 7, 2012.

C. Outreach Activities

ICPSR attends large national conferences and meetings of professional associations as well as regional conferences. The national conferences are better for learning about potential data acquisitions and providing information about the ICPSR Summer Program. Regional conferences are more likely to generate new revenue through memberships. Every two to three years, ICPSR selects and chooses three to four regional conferences to attend. ICPSR encourages staff and archives to get on the conference programs through panels and workshops.

D. ICPSR on Facebook and Twitter

ICPSR makes use of social networking tools in conjunction with webinars and e-mail for communications. Currently, there are 600 Twitter followers and 539 Facebook fans. Webinars average 80 registrants with a 70 to 80 percent show rate (the average show rate for webinars overall is 30 to 40 percent). PowerPoint files used in workshops, panels, webinars, and presentations are loaded to SlideShare to make it easier for others to discuss ICPSR's value.

ICPSR does few press releases but is trying to change that. It was suggested that we send the information to journals and their editors since they have people who will do most of the work if they have the content.

E. Resource Center of Minority Data (RCMD)

David Thomas discussed proposed new directions for RCMD, and the highlights of the RCMD re-invention plan were quickly reviewed. Committee members were asked to act as ambassadors for the RCMD and its efforts to increase the availability and usage of data about under-represented populations. Council members offered to assist by placing a link to the RCMD on their respective center and/or department sites.

Preservation and Access Committee

Council: Ann Wolpert (Acting Chair), G. Sayeed Choudhury (by phone), and Susan L. Cutter

Staff: Bryan Beecher, Nancy McGovern (by phone), and Mary Vardigan

The committee began by congratulating G. Sayeed Choudhury, who had recently been named the 2012 winner of the Frederick G. Kilgour Award for Research in Library and Information Technology. The award, jointly sponsored by OCLC (Online Computer Library Center, Inc.) and the ALA (American Library Association), is given for research relevant to the development of information technologies, especially work that shows promise of having a positive and substantive impact on any aspect(s) of the publication, storage, retrieval, and dissemination of information, or the processes by which information and data are manipulated and managed. Choudhury was recognized for his leadership in the field of data curation through the NSF-supported Data Conservancy project.

I. Strategic Plan Update and ICPSR Leadership Role

The committee reviewed a list of topics and areas that the Preservation and Access Committee has covered in recent years. The focus has generally been on preservation and access issues and the technologies necessary for the ICPSR repository to fulfill its preservation and access missions effectively.

The Strategic Plan, developed over the course of 2007-2008, has a strategic direction specifically focused on leadership for ICPSR. Council has in the past encouraged staff to the Preservation and Access Committee to take advantage of leadership opportunities as the organization is uniquely positioned to make contributions to the social science research community and beyond. The committee discussed the fact that the plan was intended to assist the organization in looking three to five years into the future and that it is probably time to update the plan. It was also pointed out that the current plan is very broad and any new document may need to provide more detailed guidance in order for the organization to take appropriate actions based on what is in and out of scope. It was further noted that this is a time of great change in the ecosystem for scholarship and data and thus a good time for ICPSR to think about how it wants to position itself.

The committee recommended that ICPSR, on the occasion of its 50th anniversary, initiate a new strategic plan. This became an action item with a “plan for the plan” required in June. The committee then turned to ICPSR’s progress in six specific leadership areas, including a gap analysis of what is needed for ICPSR to play a leadership role in the community for each of these topics.

Data Citation and persistent identifiers. There are now many players in this area and the effort has gathered significant momentum. ICPSR can play a convener role, bringing people together to push this area forward. We can also work with content aggregators like Thomson Reuters to ensure that ICPSR data are discoverable, citable, and associated with publications based on them through DOI linkages. More specifically on this latter point, ICPSR has been working with

Thomson Reuters on new services designed to link publication and data in the Web of Knowledge. Council members advised that the new functionality should give ICPSR (and other data providers) appropriate credit and appropriate branding.

Another point raised was that the business model planning committee of the ORCID (Open Researcher and Contributor ID) initiative has tasked Ithaka to contact universities and other organizations for input into the optimal financial model for ORCID. (Ithaka is a non-profit that helps the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways.) ORCID is interested in determining how its new service to provide unique identifiers for researchers might be paid for by universities and others. ICPSR should weigh in on this topic.

ICPSR is working on a proposal to the Sloan Foundation to convene a meeting of journal editors, data archives, and other stakeholders to discuss data citation policies in the social sciences. A second thread of the proposal is to convene a meeting of domain repositories to discuss data citation and other salient data-related issues that cross domains. Council is very supportive of the ICPSR convening role. It is hoped that the grant will create traction and that the challenge grants being considered as part of the first meeting will extend the impact of the conference in important ways.

Documentation standards. This has been a leadership area for ICPSR since 1995 when ICPSR established the Data Documentation Initiative with the goal of creating a structured metadata standard for the social sciences. There is a lot of work going on in this space and ICPSR was urged to continue its work in leading and supporting the DDI project. Council was supportive of the recent NSF application for enhanced DDI metadata markup for the American National Election Studies (ANES) and the General Social Survey (GSS) series.

Provenance. This is an emerging area of emphasis for ICPSR and the greater research community. ICPSR would like to be perceived as an active contributor by playing a convener role to raise awareness and stimulate action. This is an ideal time for ICPSR to work on filling the gaps through activities such as the metadata project described above. Part of the project to tag the documentation for the ANES and GSS in DDI addresses provenance in the sense that ICPSR would enhance metadata to indicate the source of individual data items and how respondents were routed through the questionnaire.

In terms of the provenance of findings and their reproducibility, the proposed Sloan-funded conference on data citation would also address best practices in providing access to the data underlying publications to facilitate replication.

Data management planning. The requirement for data management plans is a huge step forward for ICPSR and others, but there is still a great deal of work to do to educate the community about best practices for data management across the life cycle. ICPSR's convener role can be leveraged by conducting training to fill these gaps. ICPSR is also interested in participating in research on the impact and use of NSF data management plans, as the requirement has now been in effect for a year, and would like to develop a library of good data management plans that have been

reviewed positively. ICPSR has been in contact with NSF directly to discuss how the requirement has been received and implemented.

There are a cluster of critical issues around the data management plan requirements that have important implications for ICPSR. The organization has been successful with its diversified funding model, relying on both external funding from the federal government and other sponsors in addition to its core membership funding. However, ICPSR may need to rethink this business model if, as we have recently heard, NSF does not consider that making data available to ICPSR members is fully open access. Currently, data deposited by individual researchers without external funding for archiving are processed using membership dues monies and made available free to members only (although others may purchase such data for \$500 a dataset). While the ICPSR membership encompasses over 700 institutions around the world and virtually all of the major research universities in the U.S., this is not considered truly open access. How should ICPSR address this?

ICPSR can encourage researchers to put a line item for archiving into their proposal budgets to enable ICPSR to offer open access to the data. This is actually being done in some cases: ICPSR staff who respond to requests for letters of support to accompany grant applications have been suggesting that funding be allocated to archiving. However, at the proposal stage it is difficult to know what might be involved in archiving – e.g., how many files and variables there will be, how much data processing will be required, etc. To address this, instead of requesting money for archiving up front, it might make more sense for PIs to request supplements to their grants at a later stage when more about the data is known.

Another approach could involve working with member institutions to solve this problem. Institutions submitting grant proposals have responsibility for complying with the provisions of the grants awarded to their PIs and thus for making data available. ICPSR could engage with the larger research universities on the social science side to find solutions – e.g., we might think about a fee added to the membership fee that would cover the cost of curation so that ICPSR could make data available free to all. The value proposition might be that all universities have essentially read-access now but that they would all have deposit access with this fee. This could be a leadership area for ICPSR.

A tension around this is that on some campuses PIs may be under pressure to publish their data in local institutional repositories (thus using this channel rather than a domain repository to meet the requirement of making their data freely available). To address this, ICPSR could set up a procedure with one copy going to the local IR and one to ICPSR. Some IRs have limits on the amount of time for which they agree to make content available; IRs could be responsible for short-term access with ICPSR taking responsibility for long-term preservation.

We need a business model that takes advantage of the capacity we already have and offers options. A white paper could be very helpful on this issue of how ICPSR should respond to the open access issue. Also, some Council members serve on an advisory board for the NSF Social, Behavioral, and Economics (SBE) directorate and may be able to discuss the issue with NSF. There is some urgency to this so it would be good to make progress by June.

Digital preservation/lifecycle management. The lifecycle approach to managing data has gained traction across the data community, but there is still a need for education in this area. To that end, ICPSR is organizing a data science course to be held this summer. The course will follow an ICPSR dataset across the data life cycle through creation, deposit, data processing, dissemination, preservation, and reuse.

ICPSR can also demonstrate good community practice as it extends to social science through the development of FLAME, which offers a leadership opportunity to showcase the use of digital preservation standards for social science data.

Restricted data. This area was covered well in George Alter's talk the previous day.

II. FLAME update

Bryan Beecher presented a brief set of slides updating Council about the FLAME project.

FLAME stands for File-Level Archival Management Engine and is an effort to create a fully OAIS (Open Archival Information System)-compliant repository that manages content at the file level rather than the study level. ICPSR's legacy approach to data has been to handle aggregations of files as studies, and this makes it difficult to ingest and manage new types of content. In a file-oriented system, relationships between and among objects are key, so developing rules for these relationships is an important aspect of FLAME.

The FLAME project will also make it possible for ICPSR to provide access to data directly from Archival Storage – a critical OAIS requirement -- rather than creating duplicate files for dissemination purposes. FLAME will generate OAIS-compliant information packages: Submission Information Packages (SIPs), Archival Information Packages (AIPs), and Dissemination Information Packages (DIPs), as well as audit trails for all actions.

FLAME will be built one collection at a time. The first area of focus will be self-archiving, with the Publication-Related Archive as an example of a type of self-archiving. Development will proceed based on the agile software development framework, which emphasizes early prototyping.

ICPSR data are currently stored in a file system with metadata stored in an Oracle relational database. We may continue with this approach or move to a repository system like Fedora (Flexible Extensible Digital Object Repository Architecture).

Beecher blogs about FLAME at <http://techaticpsr.blogspot.com/>.

III. Update on Data Science Course

This course, titled "Applied Data Science: Managing Data for Reuse," is taking shape and will be held the week of July 23 with up to 25 participants involved. Nancy McGovern suggested such a course last June and refined the idea in October.

It makes sense for ICPSR to provide education and training in data management and curation as part of the Summer Program. The Summer Program has for several years been offering a course for data librarians and ICPSR Official Representatives (ORs) who provide data-related services called “Providing Social Science Data Services: Strategies for Design and Operation.” The new course will focus on practicing professionals who identify as data curators or have a strong interest in the field. There will also be a third course related to data management offered this summer: “Assessment and Mitigation of Disclosure Risk in Data: Essentials for Social Science.”

For the Applied Data Science course, ICPSR is partnering with the University of Michigan’s School of Information, which is providing the venue and some experienced instructors. Jake Carlson, an academic librarian at Purdue, will serve as an instructor in both this course and the Providing Services course. He has done a lot of work on Data Curation Profiles.

ICPSR is experimenting with Google AdWords to see if the paid search will draw people into the course. Council gave suggestions about other lists and venues in which to advertise the course.

IV. Update on NSF Data Management Plans

Choudhury gave an update on activities at Johns Hopkins University (JHU) surrounding data management plans and archiving data. Building upon the work of the Data Conservancy, JHU has launched JHU Data Management Services, with two components supporting NSF proposals:

- *Pre-proposal submission consultation.* This optional service provides assistance to NSF PIs in creating data management plans. There is no charge to the PIs as a group of schools and departments at JHU pay for the positions that do the consulting, based on the average volume of NSF proposals submitted annually. JHU faculty submits over 200 NSF proposals each year.
- *Data management support that includes data depositing and preservation for a specified period of time.* These post-award services will be funded initially through a percentage of award fees from each grant. PIs may request financial support for JHU archiving services in their grant proposals. About 40 percent of the PIs creating data management plans using the JHU service have agreed to pay for this. The formula for costs is a first pass and will need to be refined as faculty are asking for fixed costs. The JHU data policy states that data will be retained for five years. This is spelled out in the Memoranda of Understanding that govern the agreements.

For social science data, the JHU Data Management Services staff recommends that PIs contact ICPSR to archive the data.

Choudhury indicated that the JHU group is interested in sharing good plans also and will be in touch with ICPSR to work on this jointly.

Training and Instruction Committee

Council: Gregory N. Price (Chair), Christopher H. Achen, and John Fox
Staff: Dieter Burrell, John Garcia, Henry Heitowit, William Jacoby, and Mary McEniry
Visitor: Panuchart Bunyakiati

I. Overview of Committee Activities

Staff reported on the Training and Instruction Committee's responsibilities and activities. The Committee addresses the educational components of ICPSR's mission, reviews ongoing efforts, and makes recommendations for future initiatives.

II. Strategic Plan

Staff reported on activities related to Strategic Direction IV, Educational Strength, which have been organized under three broad categories. First, there is maintenance and expansion of the Summer Program's curriculum. Plans for the 2012 Summer Program include the traditional roster of courses, as well as a number of new courses. The latter include three new four-week classes and 13 new short courses. Innovations include three workshops on data science, two short workshops on statistical computing, a new emphasis on course sequences to provide more comprehensive coverage of relevant topics, and several workshops on topics of special interest to diverse audiences.

Second, there is a focus on broadening and enhancing the diversity of the Summer Program faculty. Staff reported that there will be 16 new instructors in 2012. Four of these are former teaching assistants or instructors. Six are women and three are African Americans. Staff also reported that longer-term efforts to diversify the Summer Program faculty have produced visible results: In 2005, there were no African American or Asian American instructors and only one Latino instructor. In 2011, there were four instructors from each of these groups.

Third, staff reported on various activities aimed at extending the reach of ICPSR's training mission. These include a new venue for Summer Program courses (Boulder, CO), new marketing and advertising efforts, partnerships with summer programs in other nations, and scholarship support for Summer Program participants provided by other institutions.

Challenges and constraints remain largely unchanged from those reported at earlier Council meetings. They include the economic climate, University of Michigan employment policies, general demands placed on junior faculty (which make it difficult for them to serve as Summer Program instructors), and the increasing costs of hiring instructors.

III. 2012 ICPSR Summer Program

Staff reported on plans for the 2012 Summer Program. There will be a total of 78 courses. Registrations are substantially higher than they were at the comparable time point in 2011.

IV. Summer Program Costs

The Committee discussed the costs of the ICPSR Summer Program. Specific topics included planning for dealing with future costs, the degree to which overall program costs impinge on establishing participant fees and staff salaries, the question of whether cost containment should be an explicit element of Summer Program operations, and whether innovations like courseware and online classes would facilitate achievement of Summer Program objectives.

V. Diversity Issues

The Committee discussed diversity issues at ICPSR in general and within the Summer Program specifically. The discussion began by considering possible strategies for obtaining financial support so that students and faculty from HBCUs could attend the Summer Program. Next, staff presented the five elements of the Proposal for Diversity Initiatives in the Summer Program: (1) appointing Hank Heitowit to the new position of Diversity Project Manager; (2) compiling a database of minority scholars who might be possible candidates for future faculty openings in the Summer Program; (3) a series of at least two minority-themed workshops in each Summer Program over the next five years; (4) presentations by minority scholars in the Hubert M. Blalock Lecture Series; and (5) establishing a working group to oversee and coordinate diversity initiatives in the Summer Program. Finally, the committee noted the existence of working groups and professional networks that already exist among young minority faculty in political science to discuss methodological issues and challenges.

VI. Community College Report

The committee discussed the Community College report, created to respond to an Action Item from the October 2012 Council Meeting.