

Needs Assessment Survey: Inter-University Consortium for Political and Social Research

Executive Summary

This Needs Assessment Survey of the Inter-University Consortium for Political and Social Research (ICPSR) was completed for the class Preserving Information at the University of Michigan School of Information. Information for the survey was collected through two visits to the ICPSR facilities, including the server room, and email correspondence with Nancy McGovern, Digital Preservation Officer.

The current state of preservation at ICPSR is very good. The organization is completing the writing of a new preservation plan and a preservation framework, both based upon the checklist for trusted digital repositories being developed by RLG/NARA. ICPSR should continue documenting their preservation strategies. They are undertaking this step now. They plan on implementing AIP storage and migration plans. Limitations primarily are a matter of getting the work done in a timely manner and updating as necessary.

I recommend they finish developing their new disaster plan as soon as possible.

The recommendations I offer are virtually all being considered by ICPSR and important to their mandate of archival preservation of their digital archives. Ultimately developing evaluation methods for their preservation plans should be a goal of ICPSR.

1. Summary of Collection

ICPSR was established in 1962 to maintain and provide access to social science data. The organization seeks to acquire and preserve data, provide open access to the data, and promote effective use of the data through teaching and training. They seek to preserve the data long-term through understanding changes in technology that warrant data migration.

The ICPSR is located in a modern building on Packard Street in Ann Arbor Michigan. The building houses their parent unit, the Institute for Social Research. A council made up of scholars and data professionals from several member universities governs the ICPSR. The organization has four major operational units: Collection Development, Collection Delivery, Educational Resources, and Data Security and Preservation. The

staff is large with most staff involved in Collection Development and Delivery, with each separate archive having its own staff. ICPSR has a Data Preservation Officer and several other staff members devoted to preservation. Two staff members work in the computing and network support services department.

Their collections are largely user driven, in that users submit their research data on the web site. ICPSR maintains over 500,000 data and documentation files from a variety of social science fields. Examples of special topics archives include National Archive of Computerized Data on Aging, the National Archive of Criminal Justice Data, Project on Human Development in Chicago Neighborhoods, and Substance Abuse and Mental Health Data Archive.

The Data Security and Preservation unit is “charged with data security planning and implementation as well as long-term archiving and preservation.” Furthermore, they protect and preserve ICPSR's “data resources in perpetuity.” The department secures back-up copies of data and documentation that are stored off-site. They also work to maintain, refresh, and migrate the files as necessary so that the information remains usable over time and thus for future researchers.

2. Preservation Planning/Policies

Present Condition, Goals, Planning, and Policies

The digital collections of ICPSR are in good condition, and the organization is planning for future preservation needs. ICPSR has hired Nancy McGovern as a Data Preservation Officer, and this has resulted in the development of a new Preservation Plan. The plan is currently being written. The document will outline the priorities of ICPSR and expand to include new kinds of digital content as needed. The plan currently being written will cover the next 3 years, and a new version will be written every two years at least.

They are also writing a digital preservation policy framework that will be different from the larger preservation plan. The framework will be based upon the Cornell Workshop on digital preservation, will adhere to trusted digital repository qualifications, and will not be updated as frequently as the larger plan, only as necessary. They plan on having a series of more specific documents for each collection. These have not been written yet.

Thus while the previous preservation plan may not adequate, the plans currently being written should be extremely helpful in guiding their preservation policy. Ideally, the organization should update and review the plan every year or at least every two years. My recommendations are to keep on the same path as currently being followed, and update the plan as frequently as necessary. The current overview and outline they have written cover everything important to preservation in a digital repository (more details discussed in section on electronic records).

Personnel, Training, and Funding

Staff members do not view the funding for preservation as adequate. ICPSR did recently receive a digital preservation grant. Staff recognizes that funding rarely goes to preservation, and by seeking out grants, they hope to show the importance of preservation to the governing council. Personnel wise, they seem to have an excellent and well-informed staff. The hiring of Nancy McGovern has likely ensured an outstanding staff for years to come.

Disaster Preparedness

ICPSR is currently working on a comprehensive disaster plan. Last year, they had a minor incident involving the cooling system in the attic above the server room, and water flowed into the server room. While all of the data was restored, this could have been a more unfortunate incident. Fortunately, ICPSR is now formulating a disaster plan. The plan hopefully will be completed in March. As stated in the outline for the disaster plan, the organization plans on conferring with the University of Michigan to work with resources already in place or available. They already have plans for a website using dplan.org, an interactive web service provided by the Northeast Document Conservation Center along with resources of the NIST Contingency Planning Guide. This is something I heartily recommend they do. Such planning will minimize future disasters.

Recommendations:

1. Continue developing the preservation plan as needed. As technology continues to change and the organization moves backups off-site, the plan should be revisited once a year if possible to see what information needs to be updated or reconsidered. ICPSR should continue evaluating their facilities to be sure they keep with the latest requirements for a trusted digital repository—they are essentially doing this already.
2. Complete the disaster plan as soon as possible. The mini-disaster in the server room suggests things can happen when you least expect it, and while ICPSR was very fortunate not to lose any data—in part because the damaged server was quickly taken to a data recovery center—if a major disaster struck the organization is in need of a well-thought out plan. For the plan, I would recommend having very specific details on emergency notification, communication between contractors and staff, clear information on data recovery, and establishing the safety of the area before saving the data. Based on my knowledge of the server room incident, there was a miscommunication between maintenance staff and the building staff, and thus the cooling system overflowed. Better communication might prevent such incidents from occurring. There should be clear communication between building staff and ICPSR staff, and the rest of the tenants of the building.

3. Building and Environment

Location and Storage

The building has been recently renovated and is sturdy. The server room is located on the fourth floor of the building, with an attic above (where the cooling system incident took

place). The server rooms are fortunately not buried in the basement. Overall, this is an acceptable environment for the servers. Moving this server room does not seem to be an option anyway. With off-site storage to maintain multiple copies of data, the server room appears adequate.

Fire and Water Monitoring/Alerts

The server room includes a sprinkler system in case of a fire, though there is concern that the sprinkler system might do damage in case of a fire. However, the damage from a major fire seems more significant than what the sprinklers might do, so this seems like a good system. In the aftermath of the server room incident (mentioned above), several changes were made to the server room. First, metal frames were constructed above the servers to prevent water from dripping down onto the servers in case of more problems with the cooling system. Second, a water monitoring system was put in on the metal frames to immediately alert staff to a problem. This seems like a good solution to the problem, and having a water monitoring system is a good idea anyway. The fire alarm system in the building is checked regularly.

Temperature and Humidity

The room includes environmental controls, including temperature and relative humidity readings. There is an alarm in the system that will alert staff if the relative humidity gets too high or low.

Storage

Data is also kept offsite, and arrangements are being worked out to store data at the San Diego SuperComputer Center. It will be supplemented by the “heterogeneous distributed storage project, an extension of the DataPASS project” which is waiting for approval by the Library of Congress. Also, the University of Michigan is planning a major data storage site, and data will be redundantly stored there as well.

Security

Security appears to be excellent. In the lobby of the building, it is necessary to check in with a receptionist before heading elsewhere. The server room requires card access, as do other entrances of the building. The server room is “an enclave”, providing secure preservation of confidential data.

Recommendations:

1. For security, fire monitoring, and humidity monitoring I have no recommendations except keep the current systems intact.
2. The plans for off-site storage are in place or coming into place soon, and I recommend they continue with these plans. If the server room were the only location their information was kept, my recommendations would be far more drastic as it is necessary to have an off-site backup for the information in case of an emergency situation. The server room itself, while adequate, is on the small side and there were concerns expressed

as to whether the monitoring system was ideal. I think it is fine; it would not be adequate without the off-site storage.

4. Electronic Records

Replacement/Reformatting

Currently their replacement schedule for their digital collections is on a ten-year plan. They hope to eventually make use of smart media, which could inform staff when hard disks are going bad. Smart media could also include servers that generate an email when they perceive trouble (heat, disc failure, etc.), media that can detect that the error rate is rising, etc. Smart media is in the future not the present. ICPSR has multiple copy management, and they always maintain a security copy. One thing pointed out is that for preservation copies of their collections, it does not have to be immediately fixed. Preservation copies are not the copies displayed online, thus you have time to fix the problem without having to rush it.

Recommendations:

1. Possibly consider having a more frequent replacement schedule than ten years.
2. Continue to investigate smart media and if feasible ultimately implement. Smart media is clearly an idea for the future that may or may not become necessary.

Trusted Digital Repository Issues

In evaluating preservation needs at an institution that primarily preserved electronic records, it is essential to take a close look at the RLG/NARA trusted digital repository checklist and see how the institution is responding to them. ICPSR plans on meeting the requirements of a trusted digital repository and is doing an excellent job of following the guidelines. I will go over each section and discuss their status. My recommendations are primarily to continue in the current direction. Obviously, the trusted digital repository checklist is itself a work in progress as the certification process is developed.

There are four major sections to the trusted digital repository checklist: organization; repository functions, processes, and procedures; designated community and intended use of information; and technology/technology infrastructure. The goal of the current standards—still in draft form—is for an institution to judge each area by planning, documentation, implementation, and evaluation.

ICPSR currently has a draft of a preservation framework. They have identified the seven sections of trusted digital repositories: OAIS compliance, administrative responsibility, organizational viability, financial sustainability, technological and procedural accountability, system security, and procedural accountability. (the above are all part of

the checklist as described above). ICPSR is addressing all of the major areas required in their framework.

A. Governance and Organizational Viability

ICPSR has a clear mission statement that is available to the public. The guidelines recommend having a contingency plan or a formal succession plan in case the digital repository ceases to exist. Currently, ICPSR is not concerned about this, as there is little chance of the ICPSR closing down. Instead, they have been sure to have backups in case of a problem at their main servers. However, having a contingency plan in case something dramatically changes at the Institute for Social Research or the University of Michigan is a good idea.

Organizational Structure and Staffing

Also as stated earlier, the ICPSR definitely has an excellent staff and a staff capable of dealing with issues as they come up and they keep up with professional development issues. All of this is addressed in the framework, and will be especially addressed in community and organizational good practice, workflow and process documents, and procedural documents.

Procedural accountability and Policy Framework

As we have seen they have plans to review the preservation plans at certain intervals. They appear to be working towards accountability with their preservation plans.

Financial sustainability

Although there may be some concern about not having enough funding, overall I would guess ICPSR is financially stable. This is an area hard to evaluate in the scope of this particular project.

Overall, ICPSR is committed and well on the way to meeting the guidelines in the checklist regarding administration and organization.

B. Repository Functions, Processes and Procedures

Ingest/acquisition of content

Material arrives at ICPSR in a wide range of formats and in various stages of completeness. The current acquisition process produces a well-formed SIP and definitions exist as to what the Submission Information Package (SIP) should look like. ICPSR staff work with the producer/depositor as needed to get all the information required. The files are normalized and the metadata edited or supplemented as needed during processing to produce the Archival Information Package (AIP)—the version that will be preserved. ICPSR thus does have the definitions necessary for SIPS and AIPS. They should continue the good work in this area to work with depositors to be sure metadata is correct.

Archival Storage: management of archived information

ICPSR is meeting the guidelines under the checklist including having definitions for each

AIP for long-term preservation, how AIPs are derived from SIPs, and verification of all information.

Preservation Planning, migration and other strategies

They have already implemented the guidelines here or at the very least are in the planning stage. The preservation framework they are writing will cover all of this information. One area I am unclear about is how they plan on evaluating their preservation planning once the new plans are fully implemented. Evaluation should be a part of the plan.

Data Management

ICPSR creates the minimum descriptive metadata necessary and ensures it is associated with the AIP.

Access Management

For access problems, ICPSR monitors what goes wrong when a user tries to access information and it fails. Records are kept and checked when needed. There is extensive background checking to monitor access requests. Most access requests go through the web site, hence that is where the checking takes place. Thus, they are meeting the requirements listed and are already at the implementation stage.

C. Designated Community and the Usability of Information

Documentation

ICPSR appears to cover the documentation necessary for designated communities through deposit forms, access policies, deposit agreements, and user agreements. In the framework, they state that “specific policies should be developed to further articulate access and use requirements and restrictions.” Thus there may need to be some clearer articulations of this information, but it is certainly planned and documented at this stage.

Descriptive Metadata appropriate to designated community

The ICPSR, as stated above, maintains minimum metadata requirements to allow designated community members to find materials of interest. They also have extensive help pages on their web site.

Use and Usability

ICPSR has developed and implemented access policies consistent with digital preservation objects, ensures that agreements are adhered to, and records actions for each object (as far as I learned during my interviews).

Verifying Understandability

I am not sure if this is implemented yet, but it is planned.

D. Technologies and Technical Infrastructure

System Infrastructure, Appropriate Technologies, and Security

ICPSR uses checksums to check integrity of the files. This system was implemented over the past 2 years. This is an excellent way to monitor AIP integrity as required in the checklist. They are currently in transition to having online copies with a failsafe offline copy of the digital content they preserve. The offline media will require a program to check an annual sample by reading it, recopying it to new media as needed. This is currently in the planning stage, but will be implemented soon. ICPSR is in the middle of this transition, and I recommend they continue on the same route.

They keep up with system security fixes as needed and keep the system current, through the systems officer. They review this periodically as needed. Anti-virus software is in place to meet processing requirements. Equipment and data are proactively monitored. They believe they are in good shape security wise, and based on my observations I agree.

As mentioned earlier, the disaster plan is still being written. They do need to test their disaster plans regularly.

Recommendations:

1. Continue working on meeting the guidelines of trusted digital repositories. ICPSR should continue documenting their preservation strategies. They are undertaking this step now. They plan on implementing AIP storage and migration plans.
2. Develop a contingency plan no matter how unlikely they will need it. This is something that they could develop while working on the disaster plan.
3. Develop a plan for evaluating their preservation plan in the future, making it easier to update and find problems faster.
4. Test Disaster plans regularly.

Summary of Recommendations/Conclusions

Short-term

1. Finish the disaster plan, preservation plan and framework, and a contingency plan.
2. Test the disaster plans at regular intervals to be sure they are prepared.

Long-term

1. Continue to meet the evolving guidelines for a trusted digital repository, and if a certification process is officially implemented, take that step.

2. Have all materials stored at several off-site locations as they are planning to do at locations in San Diego and across the country, and possibly Europe.
3. Develop a new plan as changes in technology occur. Plans need to be updated at least once a year ideally. Realistically this may not happen, but the more frequently procedures can be updated the safer the data will be.
4. Develop an evaluation plan for preservation policies that will work in conjunction with the preservation plan and framework currently being written.
5. Adopt Smart Media technology if and when feasible.