WHY PRESERVATION NETWORKS NEED COLLECTION AUDITING – AND HOW TO DO IT

Micah Altman <micah_altman@alumni.brown.edu> | Jonathan Crabtree <jonathan_crabtree@unc.edu>
Thu-Mai Christian <tlchristian@unc.edu> | Nancy McGovern <nancymcg@mit.edu> | Sophia Lafferty-Hess <slaffer@live.unc.edu>

SafeArchive: A Tool for Automated Preservation Policy Auditing

- Enables TRAC audits of distributed storage networks
- Automates collection replication based on policy
- The reliability of a top-down replication system & resiliency of P2P

Investigating Real Preservation Networks & Real Collections

Phases: 1) setup; 2) self-audit; 3) test compliance ; 4) audit PLNs

Council of Prairie and Pacific University Libraries
- 9 Institutions
- Dozens of collections – 10,000s documents, images, ETDS
- Goal - ‘Multiple’ verified replicas

Digital Federal Depository Library Program “USDocs”
- Dozens of institutions
- 580+ collections – millions of documents
- Goal - “Many” replicas, “many” regions

Data-PASS
- 6 institutions
- 5 collections – 100,000s of databases and documentation files
- Goal - 3 replicas, in at least 3 regions

Lessons Learned

- Trust, but continuously verify
- Distributed digital preservation works
- Replication agreement is not collection agreement
- Transparency aids preservation
- Distributed preservation requires distributed auditing

Findings

- Most collections were successfully replicated, however ...
- Networks unable to demonstrate policy compliance during audit

SafeArchive is a collaborative effort of the Data-PASS Partners: ICPSR <http://icpsr.umich.edu>; Roper Center <http://ropercenter.uconn.edu>; Odum Institute <http://www.irss.unc.edu/odum>; NARA <http://archives.gov>; and Institute of Quantitative Social Science (IQSS) <http://iq.harvard.edu/>. It is managed through IQSS and works in collaboration with the LOCKSS project at Stanford University. The project is sponsored by the Institute of Museum and Library Services (IMLS), under award #LG-05-09-0041-09.