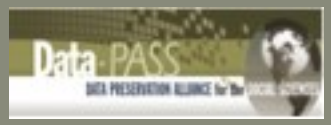
SafeArchive

Audit System Overview

Jonathan Crabtree
Odum Institute UNC Chapel Hill

Collaborators*

- Micah Altman, Leonid Andreev, Ed Bachman, Adam Buchbinder, Ken Bollen, Bryan Beecher, Steve Burling, Tom Carsey, Thu-Mai Christian, Kevin Condon, Jonathan Crabtree, Merce Crosas, Gary King, Patrick King, Sophia Lafferty-Hess, Tom Lipkis, Freeman Lo, Jared Lyle, Marc Maynard, Nancy McGovern, Lois Timms-Ferrarra, Terry Rowland, Akio Sone, Bob Treacy



- Research Support

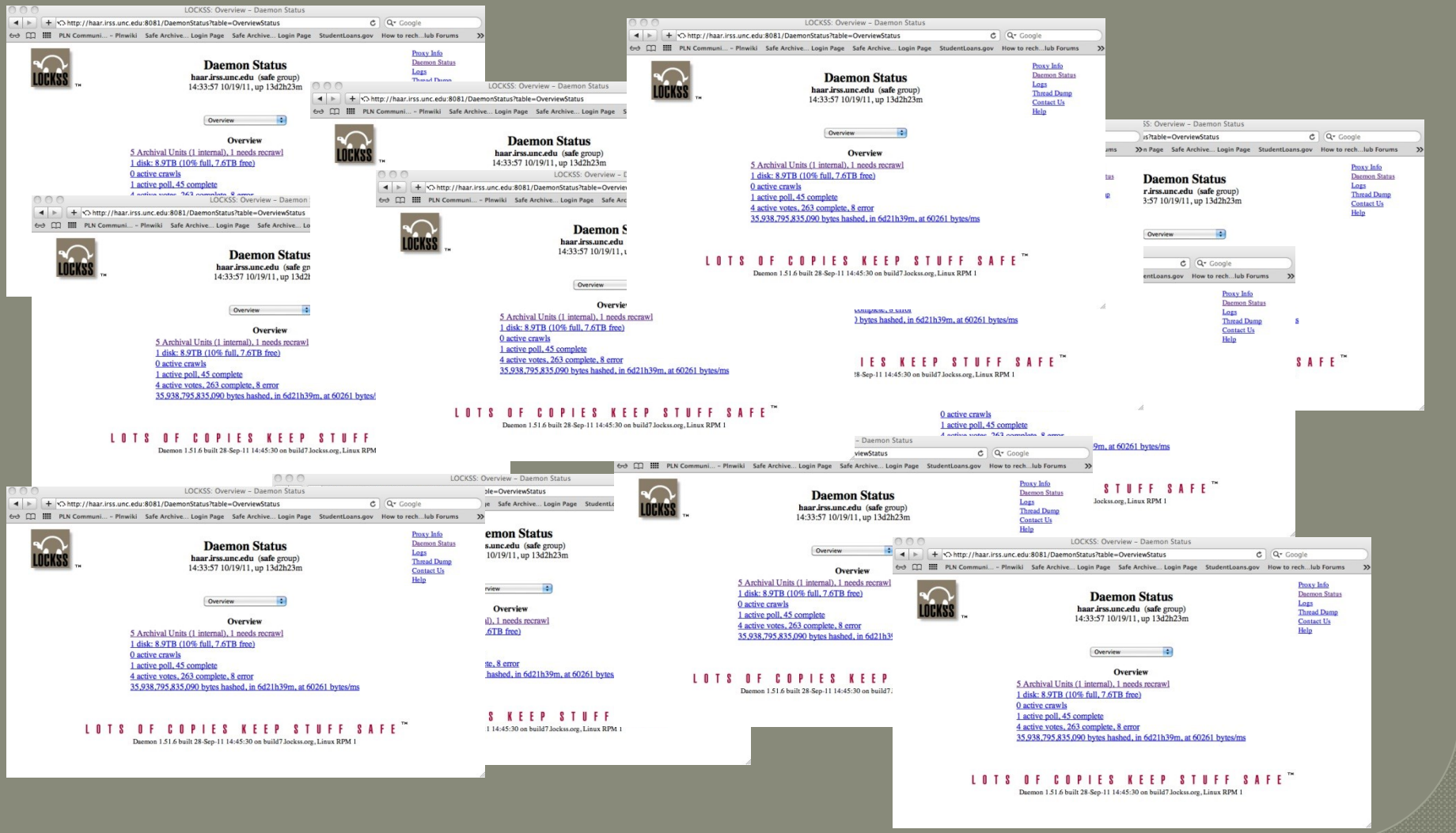
Thanks to the, IMLS (LG-05-09-0041-09), Library of Congress (PA#NDP03-1), the National Science Foundation (DMS-0835500, SES 0112072), the Harvard University Library, the Institute for Quantitative Social Science, the Harvard-MIT Data Center, and the Murray Research Archive.

* And co-conspirators

Related Work

- Reprints available from: http://futurelib.org
- Altman, M., and J. Crabtree, 2011. "Using the SafeArchive System: TRAC-Based Auditing of LOCKSS", *Proceedings of Archiving 2011*.
- Thu-mai Christian, Jonathan Crabtree, Nancy Mcgovern et al., Overview of SafeArchive : An Open-Source System for Automatic Policy-Based Collaborative Archival Replication. Proceedings of iPres 2011. (Forthcoming)
- Altman, M., Beecher, B., and Crabtree, J.; with L. Andreev, E. Bachman, A. Buchbinder, S. Burling, P. King, M. Maynard. 2009. "A Prototype Platform for Policy-Based Archival Replication." *Against the Grain*. 21(2): 44-47.
- Altman, M., Adams, M., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., & Young, C. 2009. "Digital preservation through archival collaboration: The Data Preservation Alliance for the Social Sciences." *The American Archivist*. 72(1): 169-182

Managing copies can be challenging

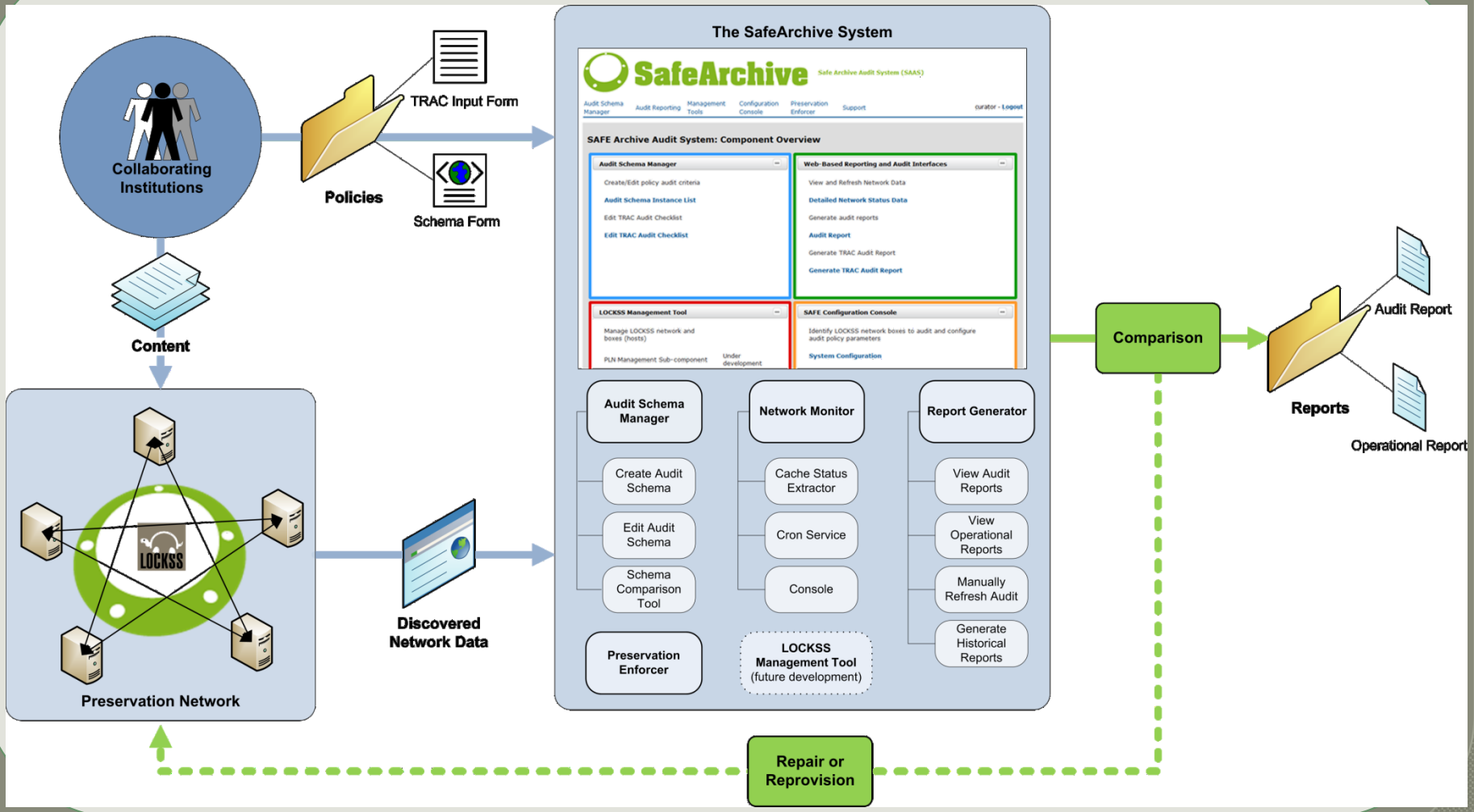


Why was SafeArchive Created?

Verified geographically-**distributed** replication of content is an essential component of any comprehensive digital preservation plan.

The requirement has emerged as a necessity for recognition and certification as a **trusted repository**.

Centralize Network Auditing



Why use SafeArchive?

SafeArchive provides the reliability of a top-down replication system with the resiliency of a peer-to-peer model.

- SafeArchive automates high-level replication and distribution policies
- SafeArchive automates multi-institutional replication
- SafeArchive facilitates sharing TRAC policies
- SafeArchive verification and audit trails for replication policies
- SafeArchive is Open Source, and integrates with *LOCKSS*, and the *Dataverse Network*
- SafeArchive is Standards-Based, and supports *DDI*, *OAI-PMH,* and *TRAC*

SafeArchive Overview

SafeArchive automates high-level replication policies and helps institutions to collaborate in preserving digital content. GUI-based tools are designed for librarians and archivists—not systems administrators.

Operationally, system users can:

- Analyze any LOCKSS network
- Check that collections are replicated, valid, and up-to-date
- Create formal replication policies
- Replicate content from web sites or digital repository systems
- Audit the network for current and historical audit compliance
- Automatically manage and repair a LOCKSS network based on a specific replication policy

SafeArchive provides the reliability of a top-down replication system with the resiliency of a peer-to-peer model.

What can you do with SafeArchive

- **Analyze** any existing set of public LOCKSS systems or Private LOCKSS Network
  - *which collections are replicated?*
  - *when were they last verified, and updated*?
  - *identify potential problems with the storage network*

- **Create** formal TRAC policies
  - *create operational policies for replication and distribution*
  - *create advisory policies for all TRAC criteria*

- **Audit** your storage network against your policies
  - *verify that collections are currently replicated, verified, updated*
  - *create historical audit trails and evidence of long-term compliance*

- **Replicate** content from web sites or digital repository systems
  - *use SafeArchive/DVN plugins to replicate content in the Dataverse Network*
  - *use SafeArchive/LOCKSS plugins to replicate content through OAI or HTTP*
  - *Automatically deploy and repair LOCKSS replication based on policy*

New Additions

- Preservation Enforcer
- New User Interface
- New Data Reports
- Track Historical Poll Agreement
- Ability to utilize agreement in less than 100% polls
- Improved Documentation
- New Amazon Instance
- New Batch Installation Scripts

SafeArchive in Action

safearchive.org

Future Possibilities

- Support additional audit standards
  - Data Seal of Approval
  - ISO 16363
- Simplify SAFE audit standards interface
  - Single interview style questionnaire "Turbo Tax Like"
  - Automatic crosswalk to all supported audit standards
- Enhance qualitative audit compliance
  - Interactive audit agency tools
  - Multiple levels of audit compliance
- Support additional replication networks
  - iRODS
  - Data Conservancy
- Audit other policy sets
  - Data Management policies
  - IRB Policies

Questions

- Website
  - www.safearchive.org
- Sourceforge
  - http://safearchive.sourceforge.net/
- Contacts
  - Micah.Altman@gmail.com
  - Jonathan_Crabtree@unc.edu