

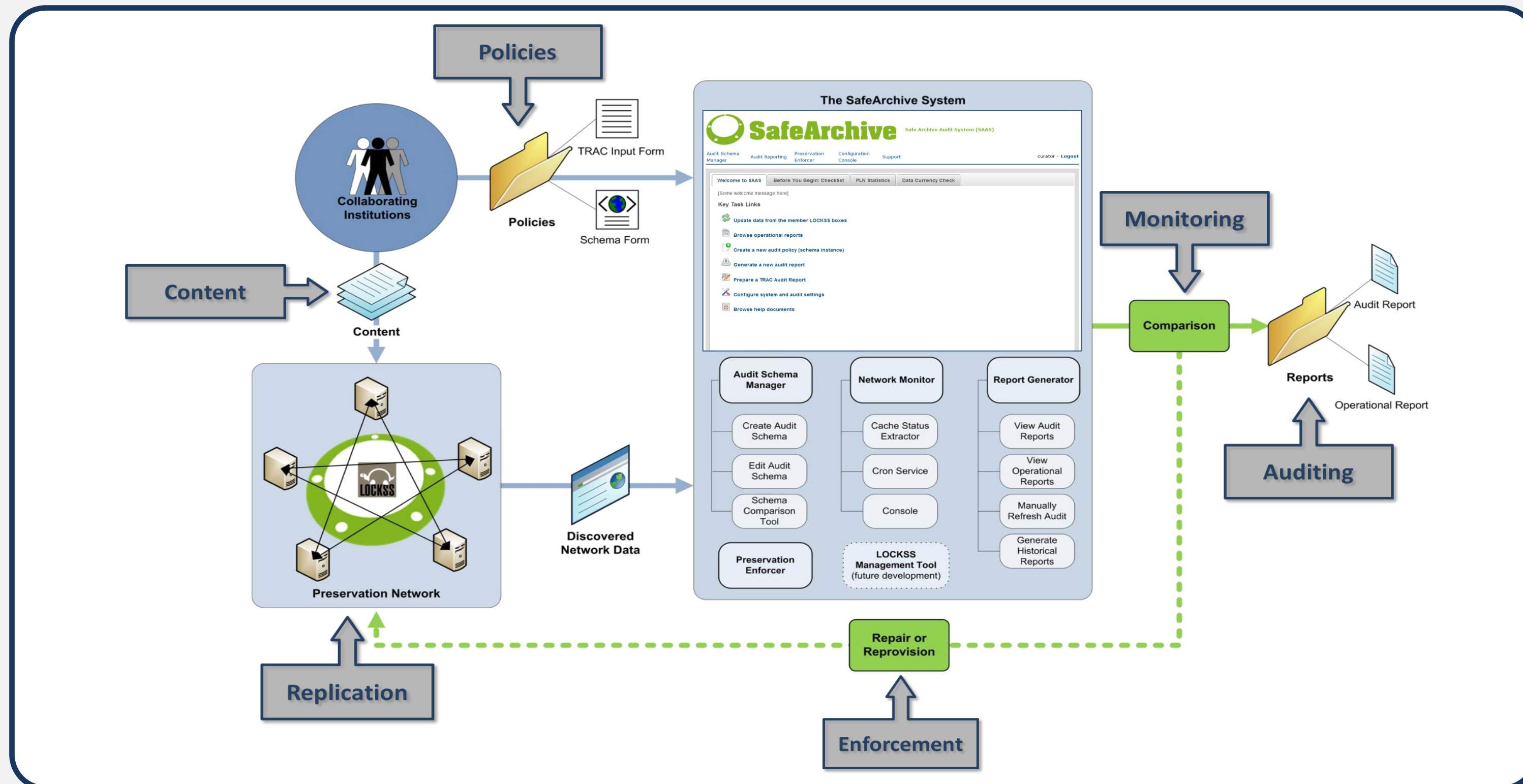
WHY PRESERVATION NETWORKS NEED COLLECTION AUDITING – AND HOW TO DO IT

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



SafeArchive: A Tool for Automated Preservation Policy Auditing

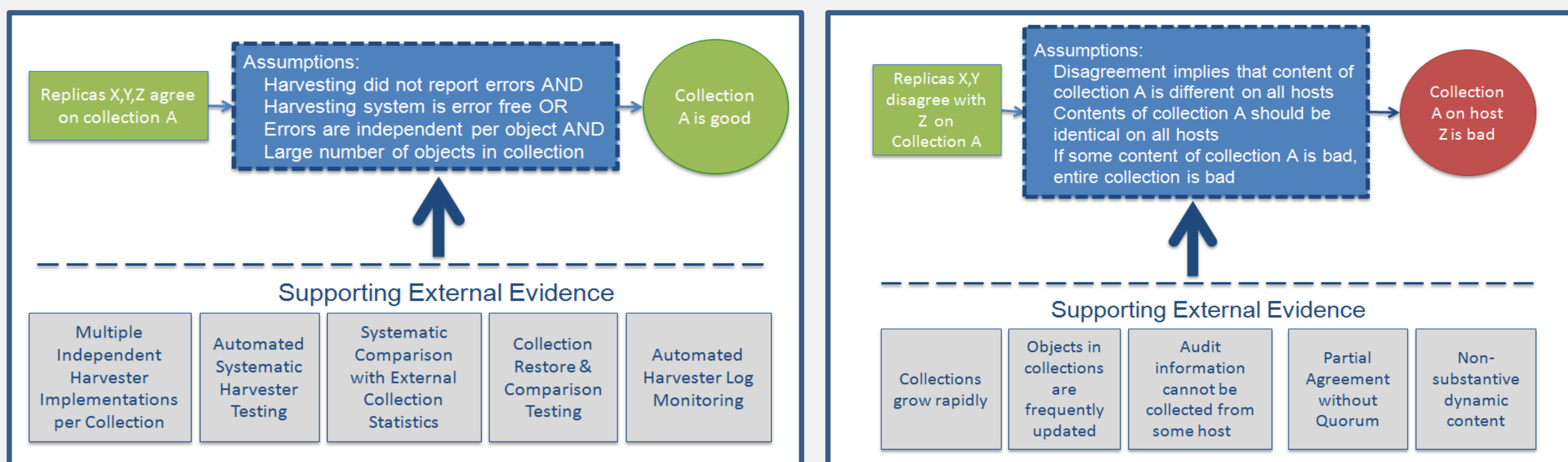
- Enables TRAC audits of distributed storage networks
- Automates collection replication based on policy
- The reliability of a top-down replication system & resiliency of P2P



New in SafeArchive Version 2.0

- More automated processes
- Trend analysis
- Easy deployment through AWS
- Many interface improvements

Trust Without Verification – What Can Go Wrong?



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

<p>Trust, but continuously verify</p> <p>--</p> <p>Things go wrong without obvious symptoms</p>	<p>Distributed digital preservation works</p> <p>--</p> <p>With appropriate monitoring and tuning</p>	<p>Replication agreement is not collection agreement</p> <p>--</p> <p>Use additional information to verify collections and policy</p>	<p>Transparency aids preservation</p> <p>--</p> <p>Visible information on operation and collections is essential</p>	<p>Distributed preservation requires distributed auditing</p> <p>--</p> <p>Standard analytics are not enough</p>
--	--	--	---	---

Findings

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

- Audit your own network
- Enable automated repairs
- Find out more
- Contact us

