

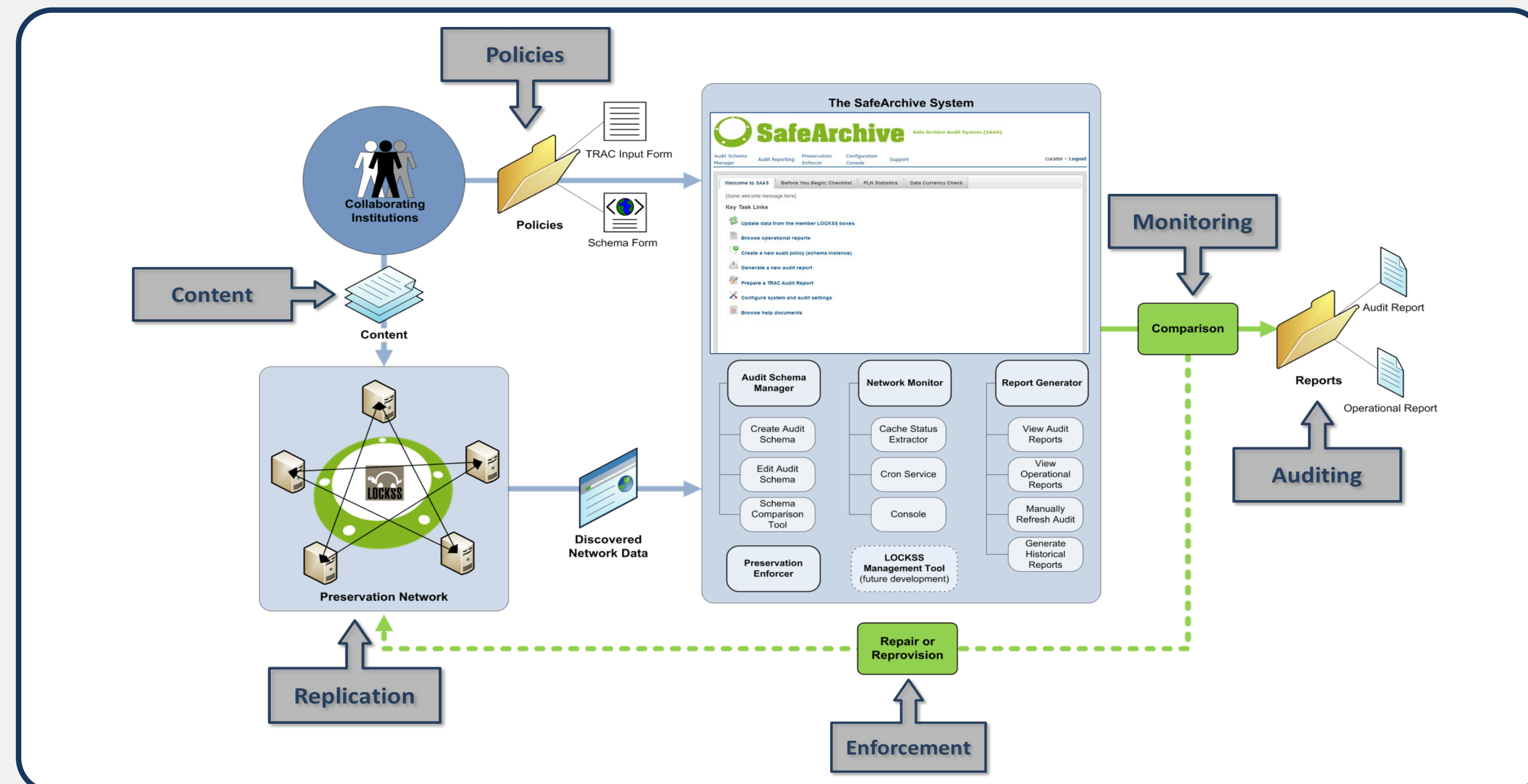
# WHY PRESERVATION NETWORKS NEED COLLECTION AUDITING – AND HOW TO DO IT

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

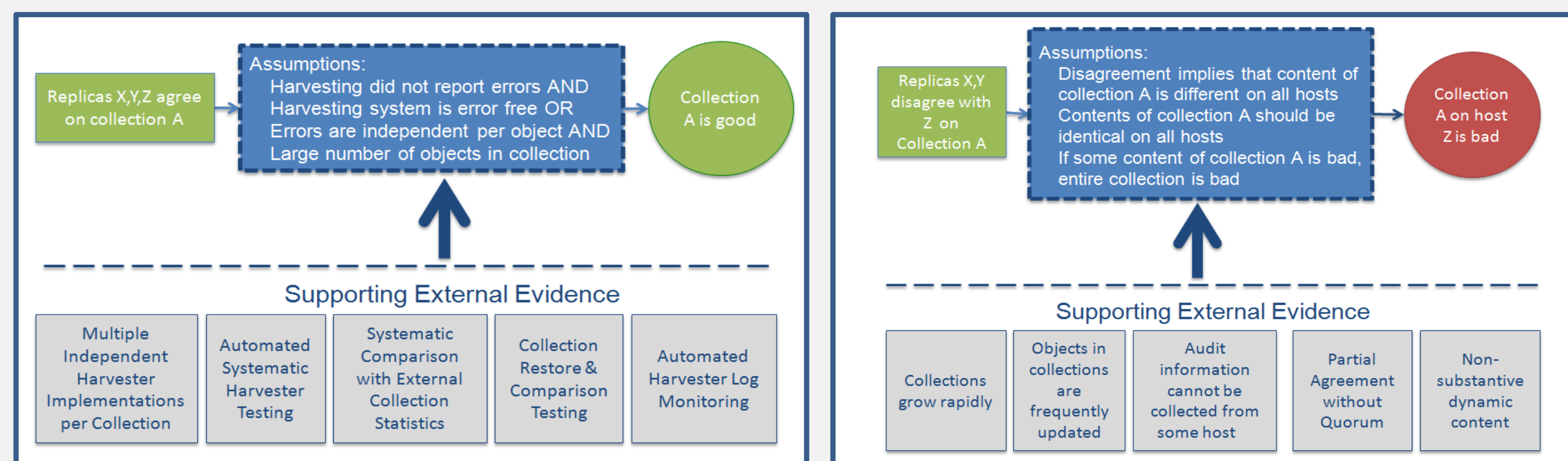
- Enables TRAC audits of distributed storage networks
- Automates collection replication based on policy
- The reliability of a top-down replication system & resiliency of P2P



## New in SafeArchive Version 2.0

- More automated processes
- Easy deployment through AWS
- Trend analysis
- 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

Trust, but continuously verify  
--  
Things go wrong without obvious symptoms

Distributed digital preservation works  
--  
With appropriate monitoring and tuning

Replication agreement is not collection agreement  
--  
Use additional information to verify collections and policy

Transparency aids preservation  
--  
Visible information on operation and collections is essential

Distributed preservation requires distributed auditing  
--  
Standard analytics are not enough

## 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

