Certifying Digital Repositories: Development of the Audit Checklist

Robin L. Dale
1 September 2006

Long-Term Repositories: Taking the Shock out of the Future
Overview

- Background & checklist development
- Uses of checklist
- “Testdriving” via the CRL–Mellon Project
- Process outputs: Compliance issues, reporting out
- Future of evaluation, audit, & certification
TDRs: Background

- **Background**
    - “…a network of certified digital archives.”
  - Open Archival Information System (OAIS)
  - *Trusted Digital Repositories: Attributes & Responsibilities*

- **Assumptions**
  - Research depends upon access digital content held by “third parties”
    - E-journals, data archives, e-science grid, etc.
  - “Wild West” of repositories and repository semantics
A semantically confusing scene

- Digital repositories
- Digital archives
- Data archives
- Institutional repositories
- Digital object management systems
- Digital asset management systems

➢ *de facto* network of varying development, capabilities, content, intent/services, data longevity, vulnerabilities, horizons/sunsets
Checklist Development

- TDR as a *system*, not just technical bits
  - Organizational infrastructure
  - Digital object management
  - Technical infrastructure
- Repositories will differ, but good practice and reliable technical infrastructure are universal
- Audit as objective evaluation; basis for trust
- Focus on transparency
  - i.e., policies, procedures, content, vulnerabilities
Uses of the Checklist

- Repository/Archives Planning
  - Set of minimal characteristics for a digital preservation repository
- Self Assessment
  - Gap analysis
- Audit
  - Organizations/Institutions
  - Services
- Basis for certification
Draft Checklist Feedback

- Too OAIS-centric
- Organizational, financial too prescriptive
- Technical too …
  - Prescriptive
  - Library-centric
  - Steeped in OAIS
- Designated Community issues
- “It doesn’t seem to apply to my archives/repository/data center/situation…”
- Too strict/too expensive – why should I do it?
- I don’t think my repository will get certified, but we certainly think others should have to…
“Testdriving” via the CRL Project

- Center for Research Libraries (CRL) Auditing & Certification of Digital Archives project
- Andrew W. Mellon Foundation funding
- Leveraged RLG-NARA checklist
- Scoped to known content
- Test audits of three archives/repositories & LOCKSS distributed archiving system
  - Multiple approaches, software, hardware, content types, state of development/production
CRL Auditing and Certification Project (2)

- Audits to determine…
  - Technologies & technical infrastructure
  - Archive processes and procedures
  - Characteristics that affect performance, accountability & business continuity

- Audit Process Development
  - Additional questions (beyond checklist)?
  - Personnel required

- Business model for certification
Lessons Learned Thus Far

- Metrics target varying granularity levels
- Audit involves far more than a checklist
  - More data needed in financial and technical areas
  - Interviews, documentation, time
- Providing evidence is critical
  - Interpretations very different
  - Existing standards only provide partial evidence
- Economics drive incentive & transparency
- Desired audit output(s) often different from completed checklist
Process Outputs: Compliance Issues & Reporting Out

- Compliance Issues
  - Is it a true checklist?
  - Levels of compliance?
  - Audit compliance versus certification grading

- Reporting Out
  - What constitutes a report?
  - Audience?
  - Who has access to outputs?
    - Information confidentiality issues
    - Community’s/depositor’s/funder’s “need to know”
    - Trusted network partners
Reporting Out: CRL Reports

- Provost/CIO statement
- Executive Summary
- Full Report
  - Introduction
    - Institutional overview
    - Scoping the audit
  - Findings
    - Organizational Analysis
      - Governance, Staff, Policies & Procedures, Financial, Succession planning
- Technical Analysis
- Content Analysis
- Vulnerabilities
- Observations & Recommendations
Checklist Rethinking & Redevelopment

- Revision based on valuable public comment, CRL test audit experience, changing landscape
- Addition of evidence, references to broader process are key
- More applicability, fewer sacrifices
  - OAIS as communication vehicle; not constraint
  - Mission/context/responsibilities, roles, content, vulnerabilities
  - Common set of policies, practices, and expectations
- User-desired outputs driving new inputs
Future of Evaluation, Audit, & Certification

- Checklist - August
- CRL project end – Oct-Nov 2006
- Products
  - Checklist
  - Supporting documentation
  - Business model to support evaluation (audit → certification; timeframe)
  - “Standards” & training
  - Economic & incentive model
  - Audit output(s)
  - Audit/evaluation registry?
Goals of Checklist, Audit & Certification

- A trustworthy system of repositories, archives, services
  - Data archives, academic content, e-journal archives, government archives, etc
- Common, objective metrics of policies, practices, and technical capabilities
- Provisions for evidence, competency, reliability, sustainability

➢ Cannot rely solely on a de facto network!
Questions?

Thank you.

Robin.Dale@rlg.org