



| D | C | C

because good research needs good data

Trust and repository audit: can repository managers assure trustworthiness?

Chris Rusbridge

June 2009



© Digital Curation Centre, 2009. Licensed under Creative Commons BY-NC-SA 2.5 Scotland:
<http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>

Funded by:
JISC



D | C | C

because good research needs good data

Outline

- Trust
- Audit
- Repository audit approaches
 - Checklist-based
 - Enhancement-based
- Preparing for audit



D | C | C

because good research needs good data

Trusted Digital Repositories





D | C | C

because good research needs good data

Trustworthy?

- Trust is directed at something by others
- Trustworthy could be:
 - Capable of being depended upon
 - Worthy of belief, as because of precision or faithfulness to an original
- Archives, in themselves and their contents, need to be trustworthy in both ways



Trust 1

- “*trust, faith, confidence, reliance, dependence*. These nouns denote a feeling of certainty that a person or thing will not fail. Trust implies depth and assurance of feeling that is often based on **inconclusive evidence**: ‘The mayor vowed to justify the trust the electorate had placed in him’”
- “Firm reliance on the integrity, ability, or character of a person or thing”
- "To trust is to **willingly relinquish control**, making yourself vulnerable to someone else for a certain outcome or consequence. Trust grows as a result of positive experiences accumulated over time." Ken Buist

The Free Dictionary



D | C | C

because good research needs good data

Trust 2

- “I cannot give them my confidence; pardon me, gentlemen, confidence is a plant of slow growth in an aged bosom: youth is the season of credulity” - William Pitt The Elder

Quotationsbook.com

- “There’s far too much trust in this room, and far too little **paranoia**” - David Rosenthal
- “Trust but verify” - Ronald Reagan & old Russian proverb



D | C | C

because good research needs good data

Trust evolution

- Trust is gained slowly, damaged easily
 - Even by events not directly related to aims & objectives (eg MPs expenses)
 - But even art library provenance fraud (eg Drewe/Myatt) substantiating forgeries doesn't completely kill trust



| D | C | C

because good research needs good data

Who must trust archives, and how?

- Funders
 - That the archive will keep resources at reasonable cost
- Depositors
 - That the archive will keep resources secure and make them appropriately available
- Consumers
 - That the resources in the archive are discoverable, usable and authentic



| D | C | C

because good research needs good data

Dimensions of repository trust

- Your repository shall:
 - Keep your objects, securely
 - Make them findable, to appropriate consumers
 - Produce them, to appropriate consumers
 - With demonstrable authenticity
 - Over time
 - At appropriate cost



D | C | C

because good research needs good data

Types of (internet) trust

- Access Trustor's resources
- Provision of services by Trustee
- **Certification** of Trustee
- Delegation
- Infrastructure Trust

Grandison, T., & Sloman, M. (2001).
A Survey of Trust in Internet Applications.
IEEE Communications Surveys and Tutorials.



Trust and certification models

- Social trust (eg selecting a builder):
 - Combination of de-centralised and centralised
 - Personal experience
 - Recommendations from those you trust
 - Holds certificate of competence
- Centralised (cf X.509)
 - Web site certificates, verified by certificate chain
 - Accredited auditors approach
- De-centralised (cf PGP)
 - Certification by a web of people, building cumulative trust
 - Eg eBay approach
 - More computational: Konfidi, Advogato, etc



D | C | C

because good research needs good data

Audit

- Formal procedure to establish confidence (usually in the **processes** of an institution), according to a defined approach
 - Eg financial audit
 - Security audit
 - Quality audit
 - Eg University QAA audits

“Audit ... tests claims made by an institution but does not compare them against a benchmark... it doesn't actually look at people teaching”

Attwood, THES 11 June 2009
quoting Philip Jones



D | C | C

because good research needs good data

Audit

- Audits rarely comprehensive
 - Usually samples transactions, policies etc
 - Should be **evidence-based**
- ISO enhancement-based approach
 - ISO 9000, ISO 27000
 - Plan-Do-Check-Act



D | C | C

because good research needs good data

Plan (establish the Information Security Management System, ISMS)	Establish ISMS policy, objectives, processes and procedures relevant to managing risk and improving information security to deliver results in accordance with an organization's overall policies and objectives.
Do (implement and operate the ISMS)	Implement and operate the ISMS policy, controls, processes and procedures.
Check (monitor and review the ISMS)	Assess and, where applicable, measure process performance against ISMS policy, objectives and practical experience and report the results to management for review.
Act (maintain and improve the ISMS)	Take corrective and preventive actions, based on the results of the internal ISMS audit and management review or other relevant information, to achieve continual improvement of the ISMS.



| D | C | C

because good research needs good data

Repository audit approaches

- Checklist-based standards
- Good practice standards
- Enhancement-based standards



| D | C | C

because good research needs good data

Checklist-based

- Most stem from Trusted Digital Repository work
 - TRAC
 - nestor catalogue of criteria
 - CCSDS RAC standardisation work
- Data Seal of Approval
 - Simplified version



Repository Audit & Certification

- Still in progress
- To lead to CCSDS & ISO standard

“4.1.3 THE REPOSITORY SHALL HAVE SPECIFICATIONS ENABLING RECOGNITION AND PARSING OF THE SIPS”

“4.2.1 THE REPOSITORY SHALL HAVE AN ASSOCIATED, PRINTABLE DEFINITION FOR EACH AIP OR CLASS OF AIPS PRESERVED BY THE REPOSITORY THAT IS ADEQUATE FOR PARSING THE AIP AND FIT FOR LONG-TERM PRESERVATION NEEDS”

etc



D | C | C

because good research needs good data

IBM Haifa tool

Long Term Data Preservation Assessment : 3.1 C. Technologies, Technical Infrastructure & Security : 3.1 C1.
System Infrastructure

client: Sample Client
respondent: Agassi, Shimon **Surveyor's Data**
line of business: Corp
engagement: LTDPA engagement
session: version 1

Long Term Data Preservation Assessment

Sub-Category name & description

3.1 C1. System Infrastructure

Navigation

Navigation Pane: Categories, Questions & status

Spider Chart

client:	Sample Client
respondent:	Agassi, Shimon
line of business:	Corp
engagement:	LTDPA engagement
session:	version 1

last updated: 18:16:36 2007 Time: 22 years 07

- ✓ Long Term Data Preservation Assessment
 - ✓ 1 A. Organizational Infrastructure
 - ✓ 1.1 A1. Governance and organizational viability
 - ✓ 1.2 A2. Organizational structure and staffing
 - ✓ 1.3 A3. Procedural accountability and policy I
 - ✓ 1.4 A4. Financial sustainability
 - ✓ 1.5 A5. Contracts, licenses, and liabilities
 - ✓ 2 B. Processes of Digital Object Management
 - ✓ 2.1 B1. Ingest: acquisition of content
 - ✓ 2.2 B2. Ingest: creation of the archivable pac
 - ✓ 2.3 B3. Preservation Planning
 - ✓ 2.4 B4. Archival storage & preservation/ mal
 - ✓ 2.5 B5. Information Management
 - ✓ 2.6 B6. Access Management
 - ✓ 3 C. Technologies, Technical Infrastructure & S
 - ✓ 3.1 C1. System Infrastructure
 - ✓ 3.2 C2. Appropriate technologies
 - ✓ 3.3 C3. Security

Contents: Licensed Materials – Property of IBM © Copyright IBM Corp. 2007 All Rights Reserved.
This document includes portions from the publication "Trustworthy Repositories Audit & Certification: Criteria & Checklist", Version 1.0, OCLC CRL, February 2007. © Copyright 2007 OCLC and CRL



D | C | C

because good research needs good data

Good practice standards

- Auditing feasible but not built-in
 - BS 5454 “Recommendations for the storage and exhibition of archival documents”
 - ISO 14721 “Open archival information systems: reference model”
 - ISO 15489 “Records management”
 - PAS 197 “Code of practice for cultural collections management”



D | C | C

because good research needs good data

Enhancement-based

- ISO 9001 “Quality management systems — Requirements”
- ISO 14001 “Environmental management systems — Requirements with guidance for use”
- ISO 27001 “Information security management systems — Requirements”
 - Information Security means “preservation of confidentiality, integrity and availability of information; in addition, other properties such as authenticity, accountability, non-repudiation and reliability can also be involved”
- DRAMBORA
 - Digital Repository Audit Method Based on Risk Assessment



D | C | C

because good research needs good data

Preparing for audit

- Be trustworthy!
 - Do what you say you will
- Be transparent
 - Keep proper records
- Aim for good practice
 - Need at least to have a position re OAIS etc
- Try a self-assessment
 - Data Seal of Approval may be a good start
 - DRAMBORA for more comprehensive approach



D | C | C

because good research needs good data

Data seal of approval



Data Seal of Approval

www.datasealofapproval.org



DRAMBORA

- Self-audit
 - Understand objectives
 - Understand staffing
 - Understand risks
- Manage risks...

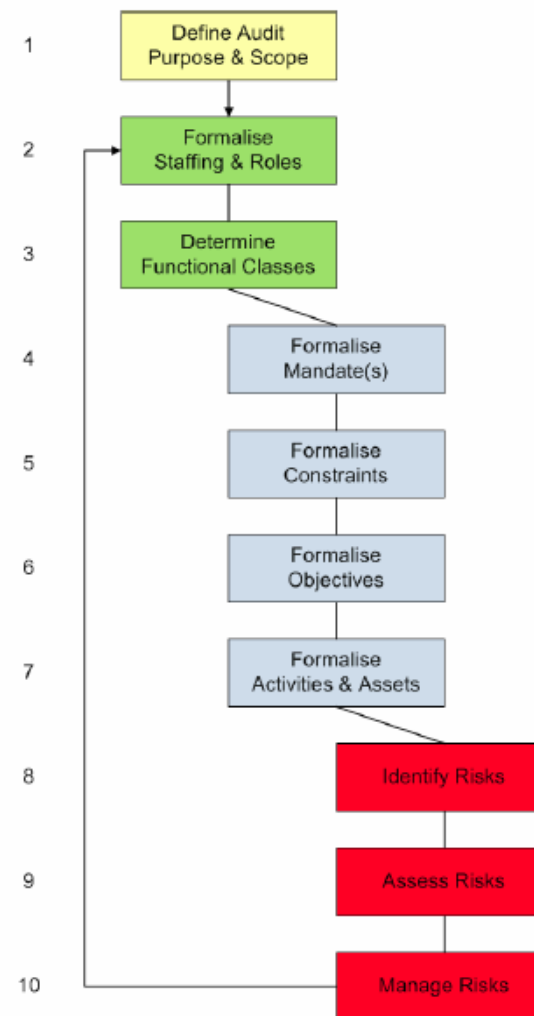


Figure 1 - DRAMBORA Workflow



| D | C | C

because good research needs good data

"Trust is the lubrication that makes it possible for organizations to work."

- Warren Bennis

But...



© Digital Curation Centre, 2009. Licensed under Creative Commons BY-NC-SA 2.5 Scotland:
<http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>

Funded by:
JISC



| D | C | C

because good research needs good data

Preservation's dirty little secret: Backup

- IT departments discouraging researchers from keeping their data on the backed-up part of their systems: volume, expense...
 - System quota 5 GB, laptop disk 150 GB
 - And... 500 GB self-powered disk < £100!
- Researchers use many different tools: Mac, Windows, Linux; desktop, laptop; on-site, off-site; corporate, self-managed etc
- No good, standardised way to organise backup
- Need standard, simple, cheap, reliable, flexible, scalable backup tools
- Job for SNIA, PASIG, Sun etc?