



IBM Haifa Research LAB

Long-Term Digital Preservation (LTDP) Assessment – from Audit to Maturity Model

Shimon Agassi, Matthew Callery, Michael Factor, Dalit Naor, Shahar Ronen
September 2008

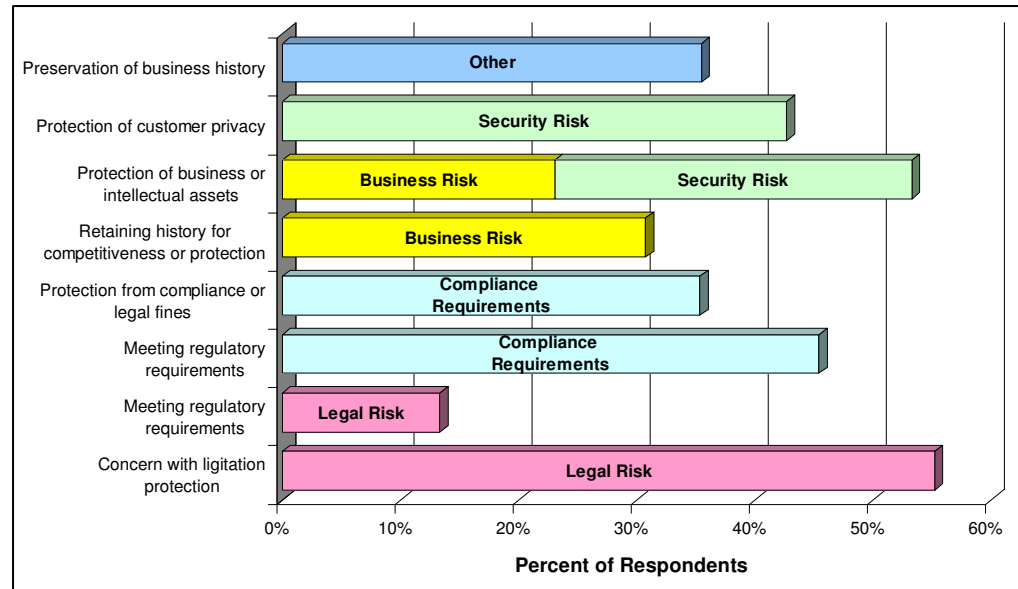
<http://www.haifa.il.ibm.com/projects/storage/ltdp/index.shtml>

| September 2008

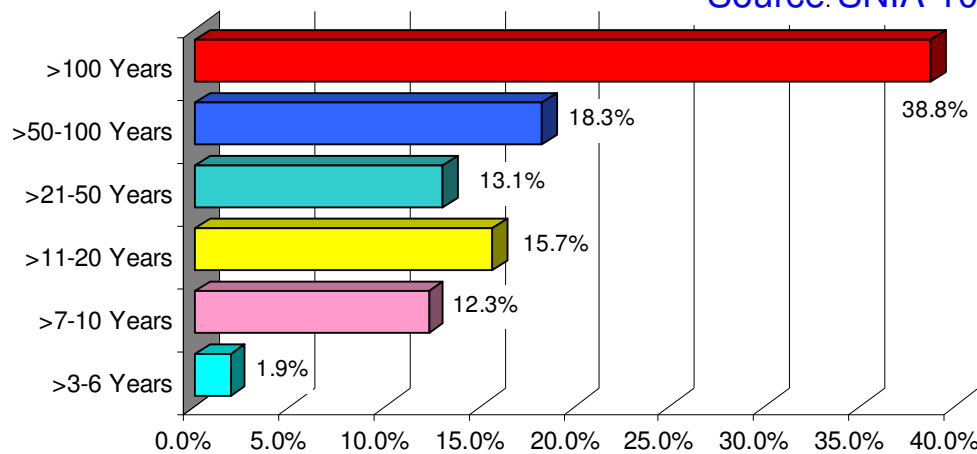
© Copyright IBM Corporation 2008

SNIA survey identified driving factors for Long-Term Digital Preservation in the Business Community

Top External Factors Driving Long-Term Retention Requirements: Legal Risk, Compliance Regulations, Business Risk, Security Risk



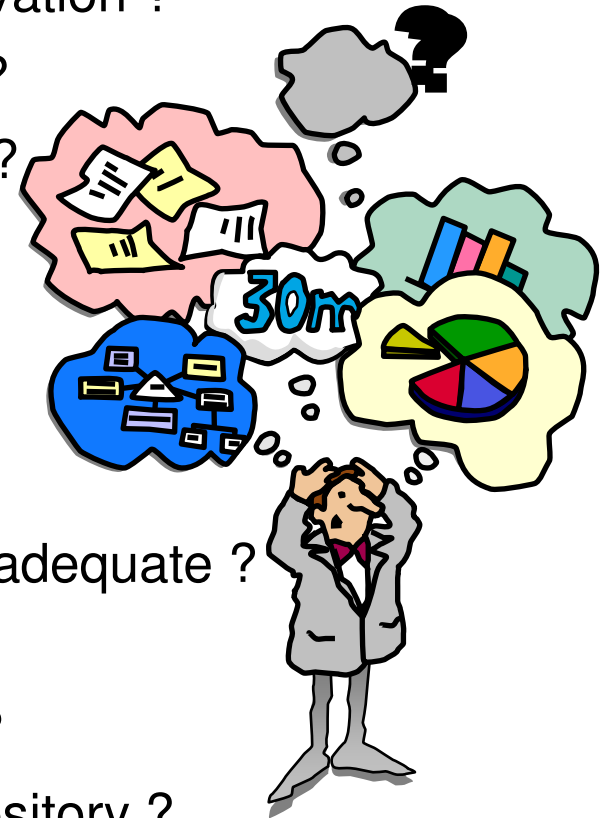
Source: SNIA-100 Year Archive Requirements Survey, January 2007.



What does Long-Term Mean? More than 20 years retention is required by 70% of polls.

Customers are facing numerous challenges of long-term digital preservation

- Is a clear Business Strategy defined for data preservation ?
- Is my designated user community clearly identified ?
- How well optimized are my preservation processes ?
- Do I need to / how well I comply with regulations ?
- How to ensure integrity and security of objects ?
- How well my IT system is capable of providing the necessary functions for long-term repository ?
- Is the metadata for description of the digital objects adequate ?
- Is the IT infrastructure adequate and scalable ?
- How much resources should I spend on repository ?
- What are my risks of not having a best practice repository ?
- Should I outsource/out-task data preservation activities ?



Online Computer Library Center, Center of Research Libraries (OCLC CRL) published the *“Trustworthy Repositories Audit & Certification: Criteria and Checklist”, Version 1.0, February 2007*

In determining trustworthiness, one must look at the entire system: *Organization, Processes and Technology:*

A. Organizational Infrastructure

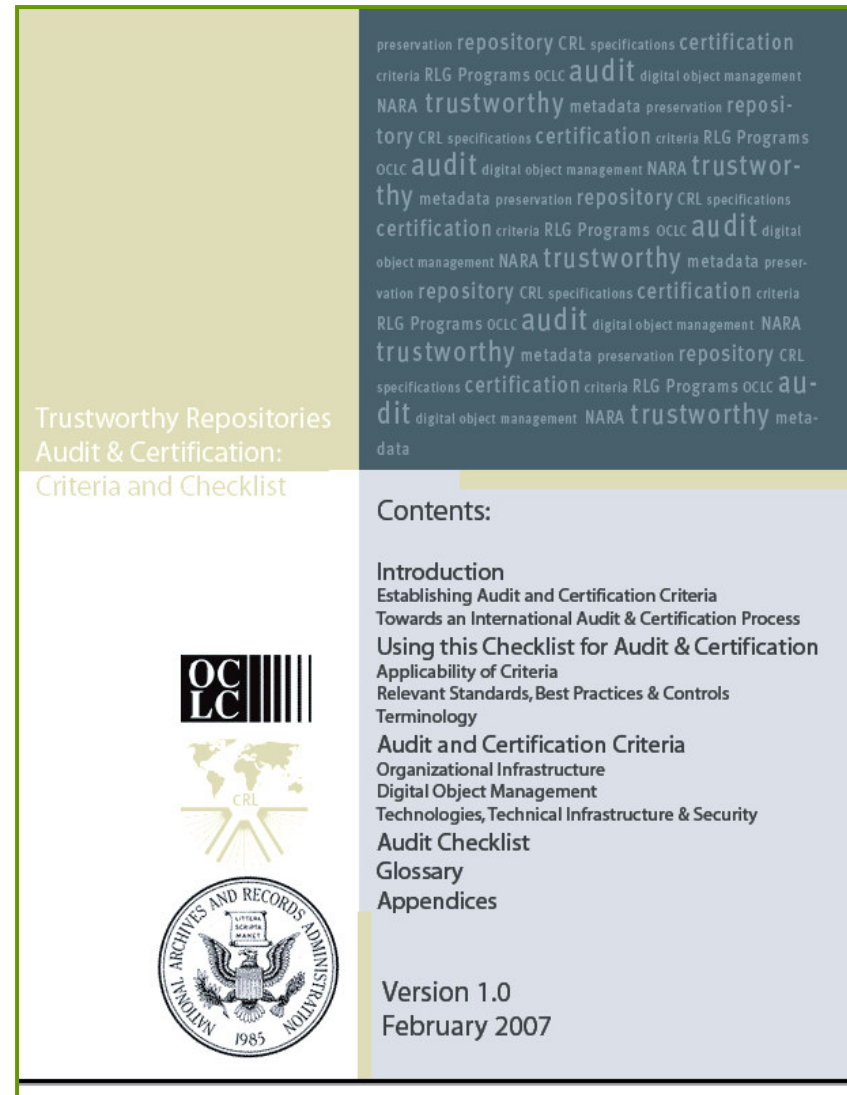
- A1. Governance & organizational viability**
- A2. Organizational structure & staffing**
- A3. Procedural accountability & policy framework**
- A4. Financial sustainability**
- A5. Contracts, licenses, & liabilities**

B. Processes of Digital Object Management

- B1. Ingest: acquisition of content.**
- B2. Ingest: creation of the archivable package.**
- B3. Preservation Planning.**
- B4. Archival storage and preservation of AIPs.**
- B5. Information Management.**
- B6. Access Management..**

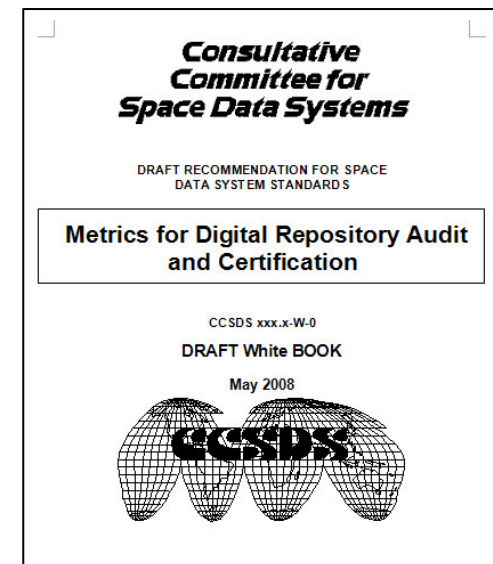
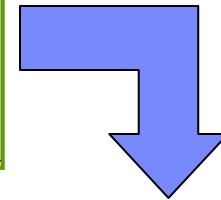
C. Technologies & Security

- C1. System infrastructure**
- C2. Appropriate technologies**
- C3. Security**



Emerging standard for Audit and Certification Standard

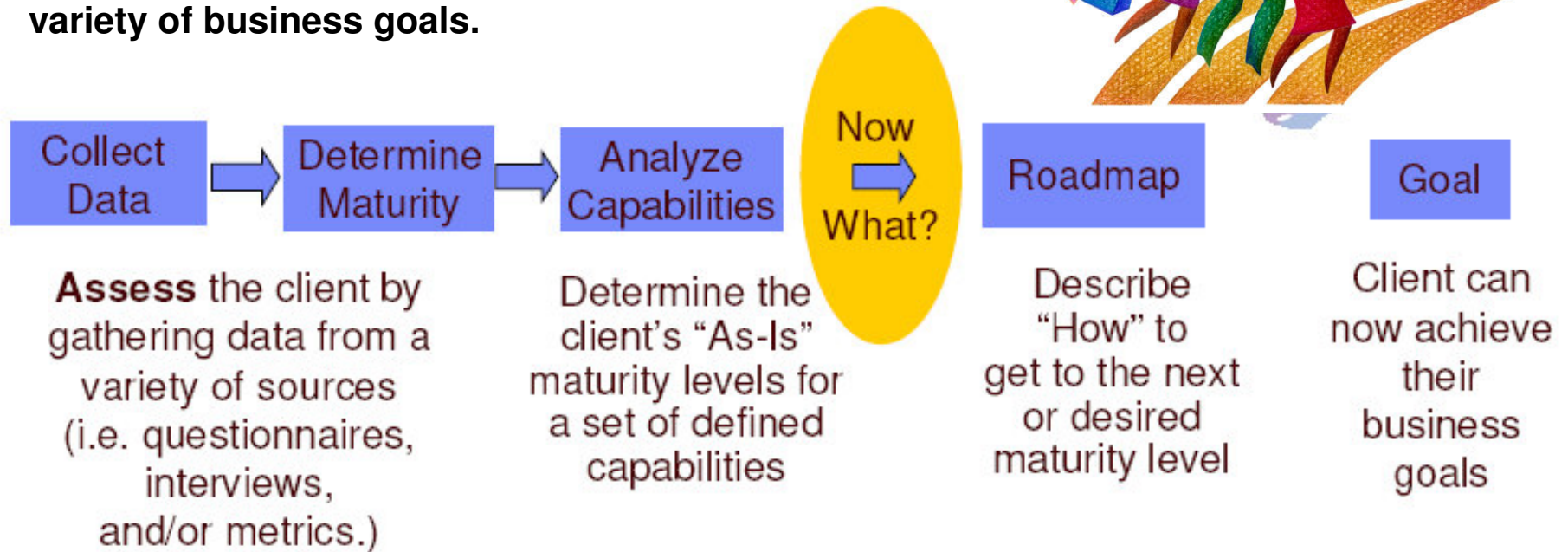
- Based on OCLC/CRL's work, an acting **Working group** is developing an ISO standard for audit and certification of trusted digital repositories.
- The aim is to develop and submit the work to ISO in the same way as the **OAIS Reference Model** (ISO 14721), namely via ISO TC20/SC13, of which the working arm is **CCSDS**.
- The work has already produced a draft proposal for the standard, called "**Metrics for Digital Repository Audit and Certification**" summarized as a CCSDS White Workbook.
- Target delivery to ISO review 3/09



Source: <http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/WebHome>

LTDP Assessment can help Organizations in identifying roadmap of actions and priorities, for bridging gaps between current and desired states

- The Audit and Certification Checklist (and emerging Standard) can serve as a base for an LTDP Maturity Model.
- An assessment based on LTDP Maturity Model can help Organizations identify gaps between Current and Desired States, and to develop a Roadmap of actions and priorities for bridging those gaps, for a large variety of business goals.



There are numerous use cases for an LTDP Assessment

- **Organizations which are planning to establish their own archive, based on reference model and best practices**
- **Organizations which already have their own archive, but want to review it against best practice, and improve accordingly**
- **Organizations which deliver archival services, and want to be ISO standard compliant, to gain advantage over competition**
- **Archival organizations/departments which seek for identifying potential risks and vulnerabilities**
- **Organizations which want to outsource their archive to service provider, and want to prepare an RFP and/or benchmark among several providers**
- **Regulatory organizations, which want to audit archival organizations (or organizations' archives) for compliance**

There are numerous potential benefits to customers from an LTDP Assessment

Identifications of gaps between current situation vs. requirements and best practice, to achieve:

- Regulations Compliance
- Help in archival services scope definition
- Trustworthiness Certification – Readiness and Audit
- Risk assessment – Business, Legal, Compliance, Security
- Potential Long-term cost savings
- Insights for optimization of design/re-design/implementation, and improvement of preservation processes and/or technological infrastructure
- Streamline interactions between producers, archives and users of information
- Help in outsourcing/in-sourcing decisions and in selection of service suppliers

LTDP Assessment tool

The LTDP tool helps in assessing the capabilities of an archival organization to deliver Long Term Digital Preservation services

Based on OCLC RLC audit checklist, the tool helps in assessing the compliance of Organization, Processes and Technology with the OAIS reference model (ISO 14721:2002) and best practices.

The Tool holds:

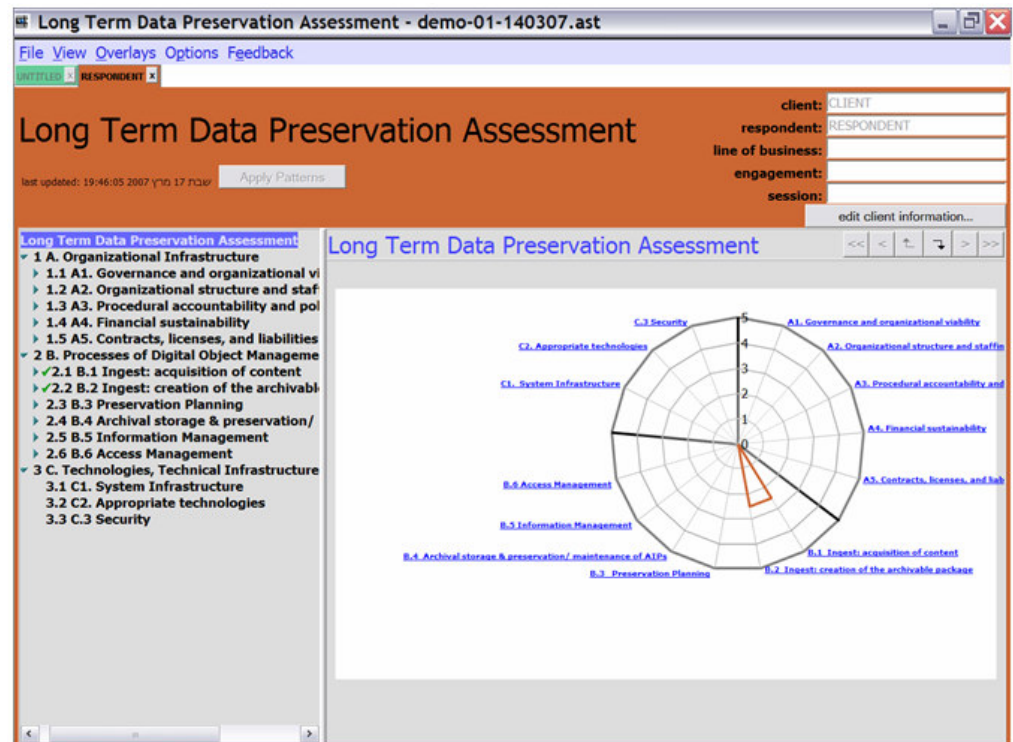
- Knowledge of expert community
- Community & ISO Best Practices
- Assessment checklist and maturity-level metrics
- Kiviat charts for comparative visualization

The Tool enables:

- Grading Current State vs. Best Practice/Desired State
- Web based Data Collection
- Numerical & Graphical analysis
- report generation
- Common repository buildup & usage

The Tool Can be used for:

- Identifying gaps between Current and Desired States
- Comparative analysis and Industry benchmarks
- Knowledge transfer

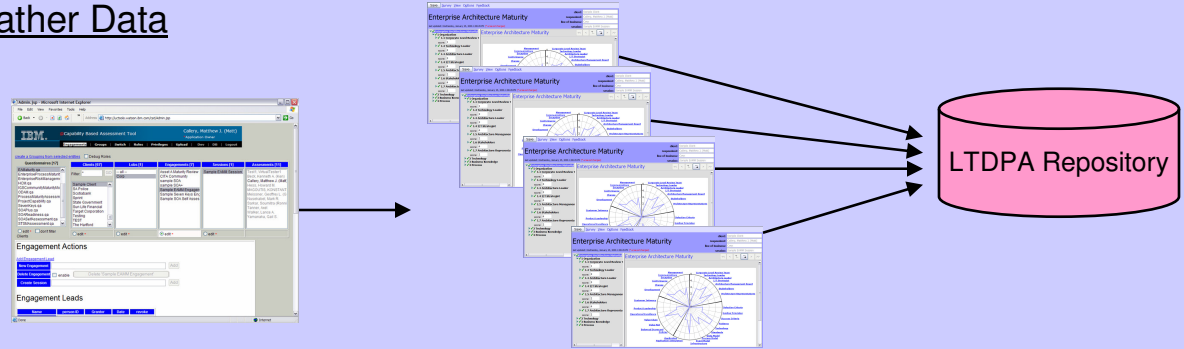


LTDPA tool usage workflow

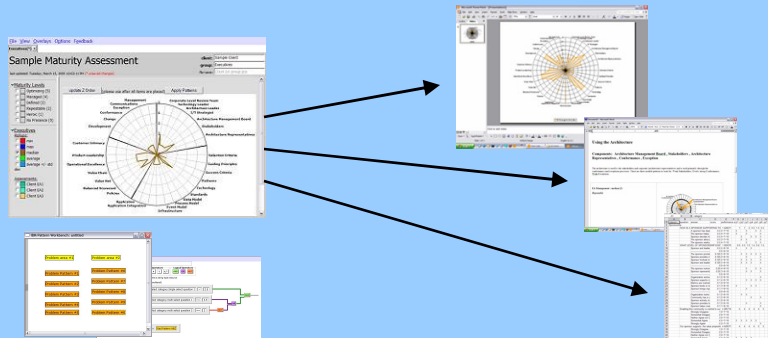
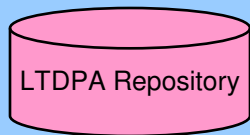
(1) Within an LTDPA engagement, engagement definition is set up, indicating individual client respondents.

The respondents can then log-in using the LTDPA web survey tool, fill out surveys and save their responses back to the repository.

1 Gather Data



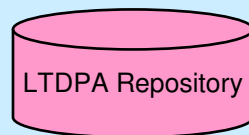
2 Analyze & Report



(2) With all of the client responses stored in the database, the engagement leader can now load the collective set of clients results, view basic statistical results, analyze the data, and export diagrams and data to MSOffice formats for deliverable creation

(3) Since all the data gets stored in the single repository, the results from a given engagement can be used again, either as a time sequence when the assessment is performed again at that same client, or as part of a benchmarking exercise, etc.

3 Historical & Benchmarking Analysis



- Year over Year
- Group vs Group
- Client vs Industry
- Etc.

Based on CAT Overview presentation, Matt Callery, IBM Research, Fall 2006

LTDP Assessment tool – Assessed criterion view

Long Term Data Preservation Assessment : 3 C. Technologies, Technical Infrastructure & Security : 3.3 C3. Security :

3.3.1 C3.1 Repository maintains regular risk assessments.

Long Term Data Preservation Assessment

last updated: 22:08:03 2007 יום שלישי 03 אפריל

client: Sample Client

respondent: Agassi, Shimon

line of business: Corp

engagement: sample LTDP

session: V

▼ Long Term Data Preservation

- ▼ 1 A. Organizational Infrastructure
- ▶ 1.1 A1. Governance and
- ▶ 1.2 A2. Organizational
- ▶ 1.3 A3. Procedural acc
- ▶ 1.4 A4. Financial sustai
- ▶ 1.5 A5. Contracts, licen
- ▼ 2 B. Processes of Digital
- ▶ 2.1 B1. Ingest: acquisit
- ▶ 2.2 B2. Ingest: creator
- ▶ 2.3 B3. Preser
- ▶ 2.4 B4. Ar
- ▶ 2.5 B5. In
- ▶ 2.6 B6. Ac
- ▼ 3 C. Technologies
- ▶ 3.1 C1. System Infrastr
- ▶ 3.2 C2. Appropriate tec
- ▼ 3.3 C3. Security
- ▼ 3.3.1 C3.1 Reposit

score: 3.6

- No Presence
- Ad hoc
- Repeatable
- Defined
- Managed
- Continually Opti

▶ ? 3.3.2 C3.2 Reposit

score:

▶ ? 3.3.3 C3.3 Reposit

score:

3.3.1 C3.1 Repository maintains regular risk assessments.

C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

Regular risk assessment should address external threats and denial of service attacks. These analyses are likely to be documented in several different places, and need not be comprehensively contained in a single document

Evidence: ISO 17799 certification; documentation describing analysis and risk assessments undertaken and their outputs; logs from environmental recorders; confirmation of successful staff vetting.

- No Presence (0):** The process does not exist or is not capable of performing as required.
- Ad hoc (1):** The process is characterized as ad hoc and occasionally even chaotic. Few activities are defined and success depends on individual effort and heroics.
- Repeatable (2):** Basic procedures are established to track functionality. The necessary process discipline is in place to repeat regular activity.
- Defined (3):** The process for both management and engineering activities is documented, standardized and integrated into a standard process for the organization.
- Managed (4):** Detailed measures of the process and product quality are collected. Both the process and products are quantitatively understood and controlled.
- Continually Optimizing (5):** Continuous process improvement is enabled by quantitative feedback from the process and from piloting innovative ideas and technologies.

clear

Comments

Observation: Although process is managed, last risk assessment was performed more than one year ago.
 Recommendation: Perform a Risk Assessment soon. IBM's Business Continuity practitioners are available for help.

Assessed criterion

Description

Metrics & Grading

Evidence

Comments

LTDP Assessment tool – General & sub-category summary view

Long Term Data Preservation Assessment : 3 C. Technologies, Technical Infrastructure & Security : 3.1 C1.
System Infrastructure

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

Long Term Data Preservation Assessment

last updated: 10:16:36 2007 אפריל 22 ראשון

- ✓ 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
 - ✓ 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 package
 - ✓ 2.3 B3. Preservation Planning
 - ✓ 2.4 B4. Archival storage & preservation/ maintenance
 - ✓ 2.5 B5. Information Management
 - ✓ 2.6 B6. Access Management
 - ✓ 3 C. Technologies, Technical Infrastructure & Security
 - ✓ 3.1 C1. System Infrastructure
 - ✓ 3.2 C2. Appropriate technologies
 - ✓ 3.3 C3. Security

3.1 C1. System Infrastructure

description

Navigation

Navigation Pane: Categories, Questions & status

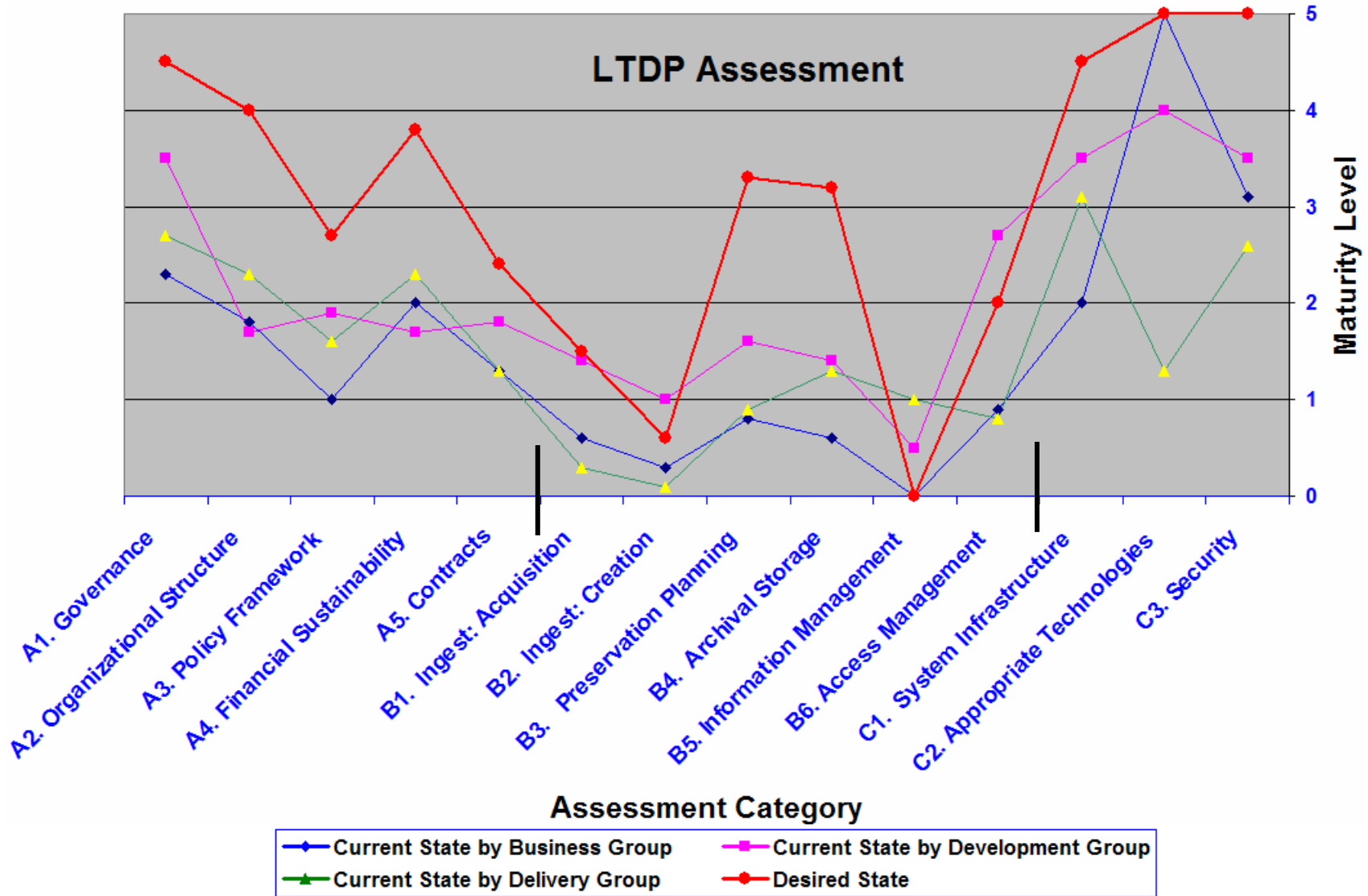
Spider Chart

Sub-Category name & description

Surveyor's Data

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

“Client X” LTDP Assessment Summary – Current-State and Desired State Maturity-Levels



More Info

<http://www.haifa.il.ibm.com/projects/storage/ltdp/index.shtml>

Michael Factor: factor@il.ibm.com

IBM Haifa Research Lab