



Digitized Newspaper Repository - Sample Assessment

The Center for Research Libraries has developed a preliminary scheme for assessing digitized newspaper repositories to objectively measure a repository's ability to provide persistent access to digitized news and to maintain the documentary integrity of that content over time. The scheme is described further in "Part 1: Digitized Newspaper Repository Assessment Scheme."

For assessment purposes, "digitized newspaper repository" refers to the platform through which a publisher makes digitized newspapers available to users and the content management and archiving system that stands behind that access platform. In this document you will find the criteria used in repository assessments as applied to "Publisher X," a hypothetical commercial content provider with a digitized newspaper platform and repository. The information included here and in the sample profile has been fabricated in order to demonstrate how the proposed assessment scheme is used to analyze and rate a digital newspaper platform and repository. Information used in the profile and assessment is based on documentation and data provided to CRL by the publisher or aggregator or information obtained independently by CRL using publicly-available sources.

All scores (for criteria component ratings, section scores and the overall composite score) are out of a possible 100 points. The score for each criterion is based on the organization's demonstrated compliance with the qualities defined by that criterion. The section scores (for Section I, II, III, IV) are compiled using the various criteria ratings, which are weighted based on their level of importance in providing long term preservation to digitized news. Finally, the composite score is derived from the four section scores, weighted based on the level of "essentialness" to responsible and transparent management of digitized newspaper repositories.



<u>Criteria</u>	<u>Source of criteria</u>	<u>Sample Assessment Information</u> <u>[Publisher X]</u>	<u>Weight (% of section score)</u>	<u>Score</u>	<u>Comments</u>
I. Organizational Infrastructure		I. Organizational Infrastructure			
I.a. Formalized organizational commitment to sustaining content integrity and longevity, providing access, and managing digital material, as well as demonstrated mechanisms for receiving and addressing feedback from designated user community	TRAC A.1, A.1.1, A3.5 ISO 16363 3.1.1, 3.3.1, & 4.6.2.1	<p>I.A. Background and Organizational Context: Publisher has 60 years of experience in licensing and reformatting newspaper holdings from collecting institutions in microfilm and digital formats while serving the research needs of academic communities</p> <p>I.B. Mission & goals: Webpage addresses institutional mission, newspaper selection and product related goals; publisher continuously solicits feedback from users and makes changes to products accordingly.</p>	30	100	
I.b. Level of specificity and formality of documented business continuity and succession plans	TRAC A1.2 ISO 16363 3.1.2.1	<p>I.C. Governance: Publisher is a division of a larger information services corporation; Publisher's administrative team includes Vice President of Product Development and Vice President of Technologies who oversee technical and intellectual aspects of newspaper products. Publisher has an agreement with a partner organization that confirms and defines conditions for succession of assets in the event of organizational downsizing or obsolescence.</p>	10	100	



<p>I.c. Evidence of staffing sufficient to support digital newspaper program/repository</p>	<p>TRAC A2.2 ISO 16363 3.2.1.2</p>	<p>I.D. Resource Allocation: Publisher has highly capable staff responsible for newspaper product development, content selection, technical support of repository systems and data support. Publisher has a staffing plan that addresses newspaper product development but plan should be revised to address, more specifically, the technical staff supporting the news products.</p>	<p>15</p>	<p>75</p>	<p>Would be improved by more explicit direction of newspaper product technology maintenance and development in staffing plans</p>
<p>I.d. Quality and comprehensiveness of business plans and strategies for securing and monitoring program financial sustainability</p>	<p>TRAC A4 ISO 16363 3.4.3</p>	<p>I.D. Resource Allocation: Publisher has a funding model that is subject to frequent fluctuations, affecting financial sustainability. However funding is actively and consistently monitored and financial planning is regularly conducted by the publisher</p>	<p>15</p>	<p>75</p>	<p>Publisher should provide proof of recent third party financial audits</p>
<p>I.e. Quality and comprehensiveness of policies that address management of legal permissions and rights and evidence of secured rights necessary for the acquisition, long term maintenance, and management of the physical and intellectual properties of ingested material</p>	<p>TRAC A5 ISO 16363 3.5.1, 3.5.1.1, 3.5.2</p>	<p>I.E. Rights Management: Publisher maintains agreements with source repositories and secures requisite rights in order to preserve and disseminate content, but is missing a consistent, formal rights management policy.</p>	<p>25</p>	<p>25</p>	<p>Publisher should create a formal, standard rights policy</p>



<p>I.f. Formalized procedures, policies and mechanisms for monitoring and addressing organizational and technological developments</p>	<p>TRAC A3.2 ISO 16363 3.3.2.1</p>	<p>I.F. Commitment to Persistence of Content: The publisher has held meetings/forums to discuss digital preservation developments with members of relevant communities and fields, but digital preservation monitoring and technology watch practices are not sufficiently addressed in the staffing responsibilities related to digital products</p>	<p>5</p>	<p>25</p>	<p>Would be improved by incorporating digital preservation related initiatives into the responsibilities of the database/repository staff</p>
<p>Section I score</p>				<p>70</p>	
<p>Total possible points</p>				<p>100</p>	
<p>Green cell indicates that criteria point is essential for successful assessment</p>					



<u>Criteria</u>	<u>Source of criteria</u>	<u>Sample Assessment Information [Publisher X]</u>	<u>Weight (% of section score)</u>	<u>Score</u>	<u>Comments</u>
II. Planning and Management of Digital Newspaper Program		II. Planning and Management of Newspaper Products			
II.a. Formalized decision making guidelines for newspaper program	Deutsche Forschungsgemeinschaft <i>DFG Practical Guidelines on digitization</i> "1.2. Selection" University of Utah Digital Newspapers <i>Project Handbook</i>	II.A. Product Implementation: Publisher has stable and publicly disclosed advisory boards to direct decisions affecting the development of news products. But publisher does not disclose minutes or reports of advisory board meetings.	30	75	
II.b. Identified program scope and documented guidelines for the selection of content	Deutsche Forschungsgemeinschaft <i>DFG Practical Guidelines on digitization</i> "1.2. Selection" Northeast Document Conservation Center, <i>Preservation Leaflets</i> "Reformatting 6.6 Preservation and Selection for Digitization"	II.A. Product Implementation: Publisher uses product gaps, authoritative historical bibliographies and user requests to guide aggregation and selection. Publisher does not disclose formal content selection guidelines	20	50	Would be improved by formal content selection policy



<p>II.c. Specifications for digitization/refinement and quality control of content data objects</p>	<p><i>ALCTS Minimum Digitization Capture Recommendations</i> <i>Library of Congress Technical Standards for Digital Conversion of Text and Graphic Materials</i></p>	<p>II.B. Digitization & Refinement: Digitization specifications are adequately described in agreements with digitization vendors in service contracts and includes file formats and metadata specifications</p>	<p>20</p>	<p>75</p>	<p>Would be improved by formal content selection policy</p>
<p>II.d. Organizational framework for implementing storage and access to repository content</p>	<p><i>Chronicles in Preservation Guidelines for Digital Newspaper Collection</i> <i>Preservation Readiness "Inventorying Digital Newspapers" & "Organizing Digital Newspaper"</i></p>	<p>II.A. Product Implementation & II.B. Digitization and Refinement: Agreements with vendors describe how the digitized material should be formatted, organized, identified and described when delivered to the publisher, including file name specifications and batch organization specifications (folder names and file structures) which is mimicked in the repository structure to facilitate ease in discovery and association of content objects</p>	<p>30</p>	<p>100</p>	
<p>Section II score</p>				<p>77.5</p>	
<p>Total possible points</p>				<p>100</p>	
<p>Green cell indicates that criteria point is essential for successful assessments</p>					



<u>Criteria</u>	<u>Source of criteria</u>	<u>Sample Assessment Information</u> <u>[Publisher X]</u>	<u>Weight (%</u> <u>of section</u> <u>score)</u>	<u>Score</u>	<u>Comments</u>
III. Content Management		III. Content Management			
III.a. Use of Open Archival Information System (OAIS) Reference Model concepts or other preservation workflow models for communicating and describing repository workflows	TRAC OAIS (ISO 14721:2002)	III. Content Management: OAIS concepts not demonstrated in publisher's disclosed documentation of its repository workflows	5	0	
III.b. Formal prioritization of digital object properties for preservation	TRAC B1.1 ISO 16363 4.1.1	III.Content Management: based on repository configuration and agreements with content sources the publisher appears to prioritize data consistency (i.e. consistent checksums throughout digital object life cycle), full text availability, content retrievability and a lossless image format	7	75	
III.c. Documented specifications for submission requirements of depositors/producers	TRAC B1.2 ISO 16363 4.1.2	II.B. Digitization & Refinement: Captured and specified in agreements with vendors and includes format, metadata and delivery specification	5	100	
III.d. Quality and comprehensiveness of ingest workflow	TRAC B2 ISO 16363 4.2	III.A. Workflow: Publisher can describe a standard ingest procedure that appears to employ data checks and responsible handling of content data objects through tools such as unique & persistent identifiers	5	75	



<p>III.e. Specifications for archiving processes (transformation and creation of archival packages); and specifications for maintaining link between archival material to source material</p>	<p>TRAC B2.3 ISO 16363 4.2.2</p>	<p>II.B. Digitization & Refinement: In order to capture transformation of material into the archival master files the Publisher requires inclusion of technical metadata describing digital conversion and migration, usually created during digitization</p>	<p>12</p>	<p>75</p>	
<p>III.f. Specifications and standards for completed archival information packages and preservation master files</p>	<p>TRAC B2.1 and B2.2 ISO 16363 4.2.1.2</p>	<p>III.A. Workflow: bi-tonal TIFF images (lossless file format) with 400 DPI resolution, with associated technical/administrative/descriptive metadata stored separately as evidenced by repository samples</p>	<p>12</p>	<p>100</p>	
<p>III.g. Procedures for verification and validation of content information and archival master files</p>	<p>TRAC B2.11 and B2.12 ISO 16363 4.2.8, 4.2.9</p>	<p>III.A. Workflow: A MD5 hash (checksum) is created for each content data object during data conversion (either by vendors or Publisher X during scanning and OCR processing), checksums are captured in administrative metadata and are used to detect bit error or rot</p>	<p>10</p>	<p>100</p>	
<p>III.h. Specifications for the description and ongoing management of content data objects</p>	<p>TRAC B2.5, B2.6, B2.7, B2.8</p>	<p>III.A. Workflow: No evidence</p>	<p>5</p>	<p>0</p>	
<p>III.i. Procedures for monitoring archival material and its storage (specifications and management of information packages, awareness of possible format obsolescence, integrity of digital material) and evidence of initiatives to maintain archival material's viability within the developing digital preservation horizon</p>	<p>TRAC B3.2 and B4.4 ISO 16363 4.4.1.2</p>	<p>III.A. Workflows: Data integrity checks built into repository workflows, and system audits are conducted on storage environments but no system for monitoring format security</p>	<p>10</p>	<p>25</p>	<p>Publisher should implement a system for monitoring possible format obsolescence</p>



III.j. Documented procedure for monitoring data corruption and loss	TRAC C1.5 ISO 16363 5.1.1.3	III.A. Workflows: Data integrity checks built into repository workflows	12	100	
III.k. Adequate metadata requirements; issue level metadata with provenance information and required title level metadata and identification as defined by designated user community (example of adequate scheme: METS/ALTO)	TRAC B5.1 and B5.2 ISO 16363 4.5.1 and 4.5.2	III.B. Metadata: Based on metadata submitted by the publisher to CRL it can be verified that metadata standards of the publisher accommodate information at the issue and title level. The publisher does maintain provenance metadata, but not always to a sufficient level as to link digital materials back to the source content	12	50	Publisher should enforce higher quality provenance information in metadata standards
III.l. Specifications for the creation and management of referential and descriptive information associated with digital objects	TRAC B5.3 and B5.4 ISO 16363 4.5.2	III.B. Metadata: Within the various file directories in the nearline and online storage units, file names capture the unique issue identifiers used to associate the production, archival and metadata files to one another	5	100	
Section III score				70.5	
Total possible points				100	
Green cell indicates that criteria point is essential for successful assessments					



<u>Criteria</u>	<u>Source of criteria</u>	<u>Sample Assessment Information</u> <u>[Publisher X]</u>	<u>Weight (% of</u> <u>section score)</u>	<u>Score</u>	<u>Comments</u>
IV. Technologies and Technical Infrastructure		IV. Technologies and Technical Infrastructure			
IV.a. Stable operating systems	TRAC C1.1	IV.A. Technical architecture: Publisher requires the creation and maintenance of systems-level technical requirements for repository operations by product and database administrators/managers to direct the development of a repository operating systems by technology staff	25	75	Would be strengthened by evidence of external support (e.g. support contracts)
IV.b. Persistent and secure backup systems	TRAC C1.2 and C1.3 ISO 16363 5.1.1.2	IV.B. Hardware, Security & Backup System: Backups in multiple, geographically dispersed areas; security verified via system security audits conducted by technology department	25	100	
IV.c. Quality and comprehensiveness of procedures for monitoring technologies and responding to changes in digital preservation	TRAC C2 ISO 16363 5.1.1, 5.1.1.1	IV.B. Hardware, Security & Backup System: Database administrator is responsible for monitoring changes in relevant technology environments. However, publisher needs better policies and practices for monitoring changes in technology and digital preservation	15	25	Publisher should develop better plans for formal technology watches and how to incorporate technology watches into workflows



<p>IV.d. Use of appropriate hardware and software technologies</p>	<p>TRAC C2 ISO 16363 5.1.1.1.1, 5.1.1.1.5</p>	<p>IV.A. Technical architecture: Nearline and online storage systems, search index, and metadata repository, all designed in house to meet the system needs as identified in systems-level technical requirements IV.B. Hardware, Security & Backup systems: Technical support staff does conduct system audits to monitor viability of technologies</p>	15	50	<p>Little information was made available regarding the specifics of the hardware employed by the repository; Technology support might be stronger if hardware/software were validated and supported by a community of developers and users</p>
<p>IV.e. Quality and comprehensiveness of procedures for testing, monitoring and updating hardware and software platforms</p>	<p>TRAC B3.3, C2.1, C2.2 ISO 16363 5.1.1.1.2, 5.1.1.1.4, 5.1.1.1.6, 5.1.1.1.8</p>	<p>IV.A. Hardware, Security & Backup systems: Publisher administers regular testing and updating of hardware and software</p>	15	50	<p>Would be strengthened by formal policy on how frequently updates and upgrades should occur with estimated/expected life cycles of hardware devices</p>
<p>IV.f. Quality and comprehensiveness of disaster preparedness and recovery plans</p>	<p>TRAC C3.4 ISO 16363 5.2.4</p>	<p>IV.B. Hardware, Security & Backup systems: None</p>	5	0	<p>No evidence of such plans; Publisher should address</p>
<p>Section IV score</p>				62.5	
<p>Total possible points</p>				100	
<p>Green cell indicates that criteria point is essential for successful assessments</p>					



<u>Sections</u>	<u>Weight (%) of section on final score</u>	<u>Score</u>
Section I. Organizational Infrastructure	40	70
Section II. Planning and Management of Digital Newspaper Program	10	77.5
Section III. Content Management	25	70.5
Section IV. Technologies and Technical Infrastructure	25	62.5
Composite Score		69
Total possible points		100