Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Measuring Latent Documentary Quality in Scholarly Communication: An Indicator Framework Based on PDF Validation and Technical Debt

Kemal Yayla

kemal.yayla@ikcu.edu.tr

Izmir Kâtip Çelebi University

Research Article

Keywords: documentary quality, latent construct measurement, scholarly communication infrastructure, technical debt, conformance validation, science metrics

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9861050/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Indicators used to evaluate scholarly communication infrastructures—openness, indexing, citation impact, and metadata availability—assess whether research outputs are reachable and recognized, but not whether the documentary objects carrying the scholarly record are archivally stable, machine-operable, and functionally accessible. This article treats documentary quality as a latent property of scholarly communication infrastructures and develops a technical-debt-based indicator framework to measure it using observable conformance validation traces. Rather than treating validation failures as isolated technical defects, the framework translates assertion-level output into a nine-family typology that classifies each failure by the impaired documentary function and the implied locus of remediation. The framework is demonstrated on a corpus of 1,340 version-of-record Portable Document Format (PDF) files from journals indexed in the Directory of Open Access Journals and associated with library publishing programs in the United States and Canada and evaluated against three successive conformance profiles for archival preservation and semantic accessibility. Results show near-universal noncompliance across all layers, but the audit reveals that nonconformance is not methodologically homogeneous. Some failures indicate patchable metadata omissions, while others reveal upstream structural deficits that cannot be corrected after document generation. Debt families exhibit discriminant behavior across disciplinary categories, supporting the construct validity of the typology as a multi-dimensional indicator system. The framework provides a scalable measurement model for transforming machine-validation output into interpretable, documentary-quality indicators, with implications for science metrics, publication ethics, and scholarly communication infrastructure research.

1 Introduction

Scholarly communication infrastructures increasingly rely on indicators of openness, visibility, indexing, citation impact, and metadata availability for evaluation. Yet these indicators often fail to assess whether the documentary object that carries the scholarly record—the version of record—is archivally stable, machine-operable, and accessible (Tabarés, 2022). A scholarly article may be open, indexed, DOI-assigned, and widely cited, yet its PDF may remain structurally opaque to assistive technologies, text-mining pipelines, and long-term preservation systems (Shanahan, 2015). This gap constitutes a measurement blind spot in science metrics: existing indicators assess whether scholarly outputs are reachable, but not whether they are documentarily functional.

The version of record (VoR) occupies a structurally distinct position within scholarly communication: it is the fixed, authoritative reference point around which citation practices, policy regimes, and academic reward structures are organized (Hinchliffe, 2022). From a document-theoretical standpoint, the VoR functions at two levels simultaneously. At the first level, it must remain archivally stable—retrievable, renderable, and self-contained over time—a condition operationalized by PDF/A (ISO 19005) (Fischer et al., 2021). At the second level, it must remain operationally usable—structurally navigable, semantically interpretable, and processable by assistive technologies, screen readers, and text-mining systems—a condition operationalized by PDF/UA (ISO 14289) (Darvishy et al., 2023; Drümmer and Erle, 2012). These

two levels are technically and analytically distinct: a document can achieve full PDF/A conformance while failing PDF/UA validation entirely (Miesenberger et al., 2024; Szentirmai et al., 2024). Prior work has shown that automated validation tools can identify many structural failures at both levels, but their raw outputs require functional interpretation before they can serve as quality indicators (Erle and Hofer, 2012; Kumar and Wang, 2024).

This article addresses this measurement blind spot by developing an audit-based technical-debt framework that translates PDF conformance validation output into interpretable indicators of latent documentary quality. The framework is demonstrated using journals associated with Library Publishing Coalition (LPC) member programs in the United States and Canada. This setting is strategically informative because it combines explicit commitments to openness, preservation, and access with operationally heterogeneous production workflows (Sandy and Mattern, 2018; Stapleton, 2019). The purpose is not to evaluate library publishing as a field, but to demonstrate how latent documentary quality can be measured within an infrastructure in which documentary stewardship is normatively salient.

This article makes three contributions. First, it conceptualizes documentary quality as a latent property of scholarly communication infrastructures not captured by conventional openness, indexing, or citation indicators. Second, it proposes a technical-debt framework to translate assertion-level PDF validation output into functionally interpretable indicators, classified by documentary function and remediation locus. Third, it demonstrates the framework on a reproducible corpus of 1,340 version-of-record PDFs, showing that preservation and accessibility failures differ not only in prevalence but also in remediation logic.

2 Conceptual background

2.1 Documentary quality as a latent property of scholarly communication

This study defines documentary quality as the extent to which a scholarly object remains technically stable, machine-operable, and functionally accessible as a carrier of the scholarly record. It is a latent property because it is not directly captured by the indicators most used to evaluate scholarly communication infrastructures. Studies of open access, scholarly visibility, and bibliographic coverage have extensively examined whether research outputs are available, indexed, discoverable, cited, or represented in metadata systems (Björk et al., 2010; Piwowar et al., 2018; Delgado-Quirós and Ortega, 2024). These indicators establish whether scholarly outputs are accessible, discoverable, or evaluatively recognized; they do not necessarily indicate whether the version of record is archivally stable, semantically structured, or machine- and assistive-system-processable.

This distinction matters because scholarly communication increasingly depends on documentary objects that circulate across heterogeneous infrastructures: publisher platforms, repositories,

bibliographic databases, indexing services, preservation systems, text-mining pipelines, and assistive technologies. A scholarly article may be open, indexed, DOI-assigned, and widely cited. At the same time, its PDF remains structurally opaque to screen readers, inaccessible to automated text processing, or insufficiently prepared for long-term preservation. Prior studies of scholarly PDFs show that formally published and openly available documents may still fail to provide the structural tagging, reading order, semantic organization, and accessibility metadata required for machine-operable access (Darvishy et al., 2023; Kumar and Wang, 2024; Nganji, 2018). In such cases, conventional indicators establish the existence and visibility of the scholarly object but not its documentary functionality.

The resulting gap is a measurement problem. Existing science metrics are generally designed to assess the circulation, recognition, or metadata representation of scholarly outputs, rather than the material and semantic properties that enable those outputs to function reliably as documentary evidence. Documentary quality, therefore, has to be inferred from observable traces in file structure, metadata declarations, encoding decisions, validation behavior, and machine-readable semantics. The task is not simply to determine whether a PDF “passes” or “fails” a technical standard, but also to translate technical evidence into interpretable indicators of the documentary functions that are impaired.

Library-based open-access publishing provides a strategically informative empirical setting for demonstrating this problem. Library publishing has emerged as a mission-oriented form of scholarly communication in which libraries support journal hosting, editorial workflows, metadata management, preservation, and open dissemination (Skinner et al., 2014; Lippincott, 2017; Colman and Markovac, 2018). This setting is especially relevant because library publishers operate under explicit stewardship commitments while managing heterogeneous production workflows, including OJS-hosted faculty journals, university press partnerships, and fully staffed publishing units (Bedford et al., 2024; Deville et al., 2019; Mattson and Friend, 2014). This combination of normative salience and operational heterogeneity makes library-based open-access publishing well-suited for testing whether documentary quality indicators can surface infrastructure-wide patterns that remain invisible to existing metrics. The purpose of this study is not to evaluate library publishing programs as a field, but to use this setting as a demonstration context for a scalable measurement framework.

2.2 The version of record as a measurable documentary object

The version of record occupies a structurally distinct position in scholarly communication. It serves as the fixed, authoritative reference point around which citation practices, policy regimes, and academic reward structures organize (Smiraglia, 2001). From a document-theoretical standpoint, this stabilizing function is not incidental. Scholarly communities require artifacts that reliably convey communicative intent across time and space, enabling shared reference, verification, and cumulative knowledge work (Buckland, 1997; Levy, 1994).

Buckland’s document theory is useful here because it shifts attention from information as abstract meaning to information as materialized evidence. Information systems do not merely circulate ideas;

they store, retrieve, transform, and mobilize documentary objects. Under this evidentiary framework, institutional designation alone does not confer documentary status. It also depends on the material and technical properties that allow an object to remain stable, reproducible, and usable across systems, platforms, and time horizons. For digital scholarly objects, these properties are instantiated in encoding decisions, metadata schemas, format specifications, tagging structures, and preservation declarations (Buckland, 1997, 2014, 2018).

This material dimension produces a two-level structure in the version of record. At the first level, the VoR functions as an archivally stable documentary object: an artifact that must remain retrievable, renderable, and self-contained over time. This dimension is operationalized by PDF/A (ISO 19005), which specifies requirements for long-term archival storage, including embedded fonts, device-independent color spaces, internal completeness, and preservation-oriented metadata declarations (Dryden, 2008; Fischer et al., 2021).

At the second level, the VoR functions as an operationally usable information object: a resource that must remain structurally navigable, semantically interpretable, and processable by screen readers, text-mining systems, and assistive technologies. PDF/UA (ISO 14289) operationalizes this dimension by setting rules for logical structure, tagging, reading order, role mapping, document language, and content–artifact distinction (Darvishy et al., 2023; Drümmer and Erle, 2012).

These two levels are analytically and technically distinct. A document can satisfy archival requirements while failing accessibility-oriented semantic requirements; conversely, some accessibility-relevant structures may be present even when a document fails a preservation profile. Prior studies of scholarly PDFs have shown that accessibility-related failures remain widespread even in formally published scientific content and that compliance with one document standard should not be assumed to imply compliance with another (Miesenberger et al., 2024; Szentirmai et al., 2024)

This distinction is not merely a technical observation about file standards; it is a measurement design principle. PDF/A and PDF/UA operationalize different documentary obligations: archival self-containment and semantic operability. Failures at each level require different forms of evidence, different interpretive categories, and different remediation logics. An indicator system that conflates them, or treats conformance status as a single binary dimension, risks mischaracterizing the nature of documentary debt.

2.3 Automated validation, semantic accessibility, and indicator validity

Automated PDF validation tools operate by checking documents against machine-readable rule sets derived from formal standards. The veraPDF validator evaluates files against PDF/A and PDF/UA profiles by testing assertion-level rules tied to specific object-level encoding decisions (McGuinness et al., 2017; Wilson et al., 2017). These outputs are technically precise but socially under-interpreted. A failed assertion indicates that a specific standard requirement is unmet. Still, it does not by itself identify which

documentary function is impaired, how severe the impairment is for scholarly use, or what kind of remediation the failure implies.

Prior work on PDF accessibility clarifies why this interpretive step is necessary. PDF/UA introduced not merely additional formatting requirements but a logical layer of document structure that connects visual presentation to semantic operability (Drümmer and Erle, 2012). Erle and Hofer (2012) further distinguish formal validity from semantic interpretability: conformance to a specification and meaningful machine interpretation are related but non-identical properties. A document may satisfy some formal requirements yet be difficult for assistive systems to interpret, or it may contain useful semantic structure yet still fail specific technical checks (Rajkumar et al., 2020). This validity–semantics distinction is central to the present study because it prevents validation researchers from treating validation output as a self-explanatory quality measure (Lehtonen et al., 2018).

Research on automated accessibility checking also shows that tool-based assessment captures only a definable subset of accessibility-relevant conditions. Automated tools are valuable because they allow scalable, reproducible, and auditable inspection of large document corpora. However, they cannot substitute for user-centered evaluation, especially where accessibility depends on interaction, task context, cognitive load, or lived user experience (Cooper et al., 2012; Rømen and Svanæs, 2012; Vollenwyder et al., 2023). Recent work on automated accessibility checking similarly emphasizes that tools vary in their coverage of guideline requirements and that machine-detectable failures should be interpreted as part of a broader validation ecology rather than as exhaustive measures of accessibility (Fischer et al., 2023).

The framework developed here, and therefore, does not claim that automated validation exhausts documentary quality or accessibility. Instead, it treats assertion-level validation output as a scalable infrastructural audit layer. The role of automated validation is to generate standardized, reproducible traces of machine-detectable documentary problems. The role of the measurement framework is to translate these traces into interpretable indicators defined by documentary function and remediation locus. When we inquire about user experience, we should complement automated validation with a user-centered assessment. Where infrastructure-scale documentary quality is the object of inquiry, automated validation provides an appropriate first-order audit layer, provided that its outputs are not read as self-sufficient indicators.

This distinction is especially important for scholarly communication research. Raw validation output is organized according to standard clauses and object-level failures; social-scientific interpretation requires categories that correspond to documentary functions. A missing PDF/A identification schema, an unembedded font, a missing document title, an absent language declaration, an untagged table, and a broken logical structure may all appear as validation failures. Yet they do not impair the same function, arise from the same workflow, or imply the same remediation pathway. Indicator validity, therefore, depends on the construction of an intermediate interpretive layer between machine-checkable rule violations and scholarly infrastructure quality.

2.4 Technical debt as a measurement translation framework

The methodological challenge in measuring latent documentary quality lies in translating low-level machine-validation output into interpretable social-scientific indicators. Raw veraPDF assertion failures—ISO clause violations tied to specific object-level encoding decisions—are not directly meaningful as quality indicators. They require functional interpretation: which documentary obligation does a given failure impair, and where in the workflow can it be remediated? This translation problem is where technical debt theory contributes not merely as a metaphor but as an indicator-construction framework.

Cunningham (1993) first introduced the concept of technical debt in software engineering to describe the long-term costs of prioritizing rapid delivery over structural correctness: decisions that accelerate initial production but defer remediation, accruing interest in the form of increased fragility, maintenance burden, and reduced adaptability. Kruchten et al. (2012) restrict the concept's analytical scope to the deferred consequences of prior technical decisions that constrain future system evolution. This restriction is productive here because it means that not all PDF nonconformance constitutes debt in the same sense. A patchable metadata omission and an upstream structural accessibility deficit represent different kinds of deferred obligation, different remediation logics, and therefore different indicators of documentary quality.

Applied to scholarly publishing infrastructure, technical debt functions as an indicator-translation framework: it converts raw validation failures into categories defined by the impaired documentary function and the implied locus of remediation. The framework performs three translations. First, from rule violation to documentary function: each failed assertion is mapped to the documentary property it compromises—archival self-containment, semantic operability, text extractability, and so on. Second, from documentary function to debt family, related violations affecting the same documentary function are grouped into one of nine families organized under two vectors—archival-preservation debt and operability-accessibility debt. Third, from debt family to remediation locus: each family is classified according to where in the production workflow the corresponding problem is most plausibly introduced and most feasibly corrected. These three translations together constitute the indicator pipeline; the output is not a conformance rate but a structured set of quality dimensions, each interpretable in terms of documentary function and actionable for remediation logic.

Preservation-related debt becomes legible when files that are archived under a compliant label fail strict validation: fonts are not embedded, color spaces are device-dependent, metadata declarations are absent—conditions that progressively undermine rendering reliability across platforms and software generations (Butler et al., 2020; Fischer et al., 2021; Lehtonen et al., 2018). Research on PDF/A conformance has shown that even documents produced with archival intent routinely miss specific metadata declaration requirements, and that these failures are often concentrated in a small number of recurring encoding decisions rather than distributed uniformly across document properties (Fischer et al., 2021, 2023). Accessibility-related debt manifests as a persistent structural deficit: the visual layer of

the VoR may be legible to human readers while the absence of structural tagging, logical reading order, and semantic role attribution renders it unprocessable by screen readers, text-mining pipelines, and assistive technologies (Darvishy, 2018; Darvishy et al., 2023; Kumar and Wang, 2024). In both cases, the underlying documentary obligation is not eliminated by deferral; it is transferred to future archivists who must remediate increasingly brittle files, to repository managers who must audit and reformat non-compliant deposits, and to users whose access to the scholarly record depends on structural properties that were never verified at the point of production.

3 Methods

3.1 Study design

This study uses a cross-sectional audit design, applying automated conformance validation to a systematically constructed corpus of version-of-record PDFs. The audit approach is suited to the study's dual purpose: generating empirically grounded compliance estimates and producing a replicable evidence base from which documentary quality indicators can be derived and interpreted. The design is explicitly oriented toward indicator construction rather than institutional evaluation: the corpus serves as a demonstration context for the measurement framework, not as a representative sample of library publishing quality. All data collections, raw outputs, and the full manifest are deposited in an open repository to support replication and extension.

3.2 Corpus construction

The study corpus was assembled through a three-stage sampling procedure designed to confine the inquiry to a well-defined institutional and quality context. In the first stage, the Library Publishing Coalition's 2024 annual directory was used to identify LPC member programs operating in the United States and Canada. The directory was parsed to extract institutional names and their associated publisher name variants as they appear in third-party bibliographic databases, yielding 112 active programs across both countries.

In the second stage, these programs were cross-referenced against the DOAJ using a publisher-name disambiguation table that accounted for variant spellings and subsidiary imprints. DOAJ inclusion served as an external quality filter, ensuring that journals in the corpus met a minimum threshold for open-access compliance, editorial transparency, and metadata completeness. This cross-reference identified 286 journals indexed in DOAJ and attributable to LPC member programs.

In the third stage, journal-level records were filtered to retain only those with at least one article published between 2024 and 2025, using DOAJ metadata. This ensured that the corpus reflected current production practices rather than legacy content. The resulting target list comprised 145 journals submitted to the VoR retrieval pipeline.

3.3 VoR retrieval pipeline

Version-of-record PDFs were identified and retrieved using a two-phase pipeline built on the OpenAlex REST API. In the harvest phase, each journal's ISSN was resolved to an OpenAlex source identifier, and the works endpoint was queried to retrieve records with a confirmed published-version open-access location within the 2024–2026 publication window, up to 15 records per journal, sorted by publication date in descending order. Each retrieved record was written to an immutable manifest that captured the work identifier, DOI, publication date, confirmed PDF URL, OA status, and a provenance timestamp. In the download phase, each manifest record with a non-null PDF URL was retrieved and verified against the PDF magic byte signature before being written to disk. Failures in URL resolution, HTTP retrieval, or signature validation were logged to a machine-readable failure file with a typed reason code.

3.4 Conformance validation and debt family mapping

The analysis proceeds in four complementary layers, which are designed to translate technical conformance outcomes into interpretable, documentary-quality indicators.

The first layer sets conformance rates through a three-step process aligned with the study's theoretical framework. Conformance was assessed against PDF/A-1b as a legacy archival baseline, PDF/A-2u as a born-digital preservation floor, and PDF/UA-1 as the operational accessibility threshold. These profiles are treated not as independent benchmarks but as a layered sequence, each representing a distinct level of documentary obligation. PDF/A-1b specifies requirements for basic archival self-containment—embedded fonts, device-independent color spaces, and internal resource completeness (Dryden, 2008; Fanning, 2017). PDF/A-2u extends this to born-digital preservation requirements, including XMP metadata conformance declarations that enable automated archival identification. PDF/UA-1 operationalizes a third, analytically distinct obligation—semantic operability—by requiring that documents contain structural tagging, a logical reading order, and semantic role attribution necessary for processing by assistive technologies and text-mining systems (Drümmer and Erle, 2012). The difference between PDF/A-1b and PDF/A-2u indicates how many documents satisfy a legacy archival standard while failing contemporary born-digital preservation requirements. The difference between PDF/A-2u and PDF/UA-1 indicates how many documents achieve preservation conformance while failing the semantic operability threshold. Conformance rates are reported at both the corpus and journal (ISSN) levels.

The second layer identifies the forms of documentary debt present in nonconforming documents. Rather than relying on raw violation counts alone, veraPDF assertion-level output was mapped to a curated nine-family typology developed for this study, in which each family is defined by the documentary function it impairs rather than by the ISO clause it violates. The typology was constructed through automated rule-to-family mapping based on veraPDF validation profiles (McGuinness et al., 2017; Wilson et al., 2017), supplemented by manual review of edge cases in which the same PDF object type carried different functional implications across clauses. The typology was informed by prior work distinguishing PDF validity from semantic accessibility (Erle and Hofer, 2012), automated compliance checking from preservation reliability (Fischer et al., 2021, 2023), and structural conformance from long-term documentary operability (Lehtonen et al., 2018; Rajkumar et al., 2020). As shown in Table 1, the nine

families are organized into two broader debt vectors—archival-preservation and operability-accessibility—each linked to a specific documentary function and remediation locus. For each document and profile, the analysis records debt family activation (binary presence), breadth (the number of distinct rule violations within the family), and intensity (the total number of failed assertions). Activation serves as the primary analytical unit; breadth as a secondary indicator of whether a problem is isolated or internally distributed; and intensity as a tertiary indicator, since some families generate repeated assertions by design rather than by severity alone.

Table 1
Debt families, documentary functions, and remediation loci

Family	Vector	Documentary function impaired	Remediation locus
Provenance & Metadata	Archival-preservation	Document identity, internal verifiability, and catalogability	Posthoc-patchable
Structural & Semantic	Operability-accessibility	Logical tag structure, reading order, content/artifact classification	Upstream-only
Table Semantics	Operability-accessibility	Machine-interpretable table hierarchy (TH/TD/THead/TFoot relations)	Upstream-only
Text Extractability & Unicode	Operability-accessibility	Copyable, searchable, machine-parsable text	Pipeline-fixable
Language Discoverability	Operability-accessibility	Machine identification of natural language for assistive technologies and NLP	Pipeline-fixable
Renderability & Resource	Archival-preservation	Reliable rendering across software and time; resource dictionary completeness	Pipeline-fixable
Color & Output Profile	Archival-preservation	Long-term visual reproducibility via ICC profile and output intent declarations	Pipeline-fixable
File Integrity & Packaging	Archival-preservation	Header, trailer, and cross-reference structure; basic archival packaging	Pipeline-fixable
Preservation Safety	Archival-preservation	Encryption, embedded files, and dynamic action restrictions	Pipeline-fixable

The third layer relates debt family activation to the remediation locus. As indicated in Table 1, each family was assigned to the workflow stage at which the corresponding problem is most plausibly introduced and most feasibly corrected. This classification serves a specific analytical purpose: it distinguishes deficiencies that require upstream intervention—and therefore cannot be meaningfully addressed after PDF generation—from those amenable to pipeline-level or post-processing remediation. Families such as Structural & Semantic Debt and Table Semantics Debt are classified as upstream-only because they depend on authoring-stage decisions concerning tag structure, reading order, and semantic hierarchy that are not reliably recoverable from finished PDFs (Pradhan et al., 2022; Rajkumar et al., 2020). By contrast, families such as Text Extractability & Unicode Debt, Language Discoverability

Debt, and several preservation-oriented families are classified as pipeline-fixable, meaning they may be mitigated through export settings, metadata correction, or post-processing scripts without altering author-facing workflows. This distinction is central to the indicator logic: it ensures that the typology produces differentially interpretable outputs rather than collapsing all nonconformance into a single undifferentiated failure state.

The fourth layer examines whether debt is evenly distributed across the corpus or clustered along disciplinary and institutional lines. DOAJ subject classifications were used to group journals into broad disciplinary categories and to compare debt family activation rates across these categories. Institutional comparisons used the journal as the primary unit of analysis and the publishing institution as the aggregation level. Because a small number of institutions contribute a disproportionate share of corpus documents, a robustness check was conducted by recomputing aggregate rates with and without the highest-contributing publishers to assess whether corpus-level patterns reflect genuine infrastructure-wide tendencies or are artifacts of output concentration.

3.5 Construct validity and indicator interpretation

Documentary quality, as operationalized in this study, is a latent construct: it cannot be directly observed but must be inferred from measurable properties of the documentary object. Two observable dimensions are distinguished. The first is preservation viability: the degree to which a document satisfies the encoding requirements for long-term archival stability, as specified in the PDF/A profile sequence (Dryden, 2008; Fanning, 2017; Fischer et al., 2021). The second is operational accessibility: the degree to which a document contains the semantic structures required for machine-readable processing, navigation, and assistive-technology compatibility, as specified in PDF/UA (Darvishy et al., 2023; Kumar and Wang, 2024). These dimensions are correlated—both reflect decisions made during document production—but are not reducible to one another. Prior research on scholarly PDF accessibility has shown that preservation conformance and accessibility conformance are independent properties that require independent assessment (Miesenberger et al., 2024; Szentirmai et al., 2024). A composite pass/fail indicator would obscure this independence by aggregating two analytically distinct obligations into a single measurement state.

Debt family activation is the primary indicator unit rather than binary conformance (pass/fail) for two reasons. First, a single failed validation may involve dozens of assertion-level violations or only one; collapsing this variation into a binary result discards information directly relevant to remediation prioritization. Second, different debt families within the same conformance profile carry different interpretive meanings: Structural & Semantic Debt under PDF/UA-1 indicates upstream authoring-stage problems that cannot be corrected without regenerating the document from source; Provenance & Metadata Debt under PDF/A-2u indicates a missing pipeline-level XMP declaration that can be patched post hoc. Binary conformance conflates these into a single undifferentiated failure state, making it impossible to distinguish infrastructure-wide structural problems from locally patchable encoding omissions.

The breadth indicator captures whether a problem is structurally isolated or internally distributed across the document by counting the distinct rule violations within a debt family. Intensity—the total count of failed assertions within a family—is a tertiary indicator, since some families generate repeated assertions by design: a document with 200 pages lacking a language attribute will produce 200 Language Discoverability assertions, a result that reflects document length as much as problem severity. Remediation locus assignments are interpretive classifications derived from the technical literature on PDF generation, digital preservation, and accessibility remediation (Fischer et al., 2021, 2023; Lehtonen et al., 2018; Pradhan et al., 2022; Rajkumar et al., 2020); they were not empirically validated against the producing workflows of journals in the corpus. The locus assignments indicate where in a generalized PDF production pipeline a given class of problem is most plausibly introduced and most feasibly corrected, based on known format-level constraints rather than on observed workflow configurations. Platform architecture may shift these boundaries in practice: XML-first authoring environments or automated post-processing pipelines may provide remediation pathways that the classification does not anticipate, such that what is classified here as upstream-only could in some configurations be addressable at the platform layer without altering author-facing workflows. This interpretive classification is therefore an analytical orientation rather than a prescriptive workflow claim—a distinction further acknowledged in the Limitations section.

4 Findings

4.1 Layered conformance as documentary quality measurement

The three-stage conformance sequence operationalizes three analytically distinct levels of documentary obligation. The first—PDF/A-1b—establishes whether a document meets the minimum requirements for legacy archival self-containment. The second—PDF/A-2u—tests whether a document satisfies the born-digital preservation floor, including the XMP metadata declarations that enable automated archival identification. The third—PDF/UA-1—tests a fundamentally different obligation: whether the document contains the semantic structures required for machine-readable processing, navigation, and assistive-technology compatibility. The gap between any two adjacent layers in this sequence is more than a compliance gap; it indicates which documentary function is systematically unmet.

Corpus-level conformance is extremely limited across all three layers. Of the 1,340 documents in the corpus, 18 (1.3%) pass PDF/A-1b, none pass PDF/A-2u, and 10 (0.7%) pass PDF/UA-1. As shown in Table 2, conformance is rare at every stage of the sequence and entirely absent at the born-digital preservation floor.

Table 2
Three-stage conformance outcomes across the corpus

Conformance profile	Role	Pass	N	Pass rate (%)
PDF/A-1b	Legacy archival baseline	18	1340	1.3
PDF/A-2u	Born-digital preservation floor	0	1340	0.0
PDF/UA-1	Operational accessibility threshold	10	1340	0.7

The gap between PDF/A-1b and PDF/A-2u is small in absolute terms but analytically significant. All 18 documents that pass PDF/A-1b fail PDF/A-2u. This loss does not reflect a broad decline in document quality across profiles; it is driven by a single recurring requirement—the absence of the PDF/A identification schema declaration in the XMP metadata block. Because PDF/A-2u treats this declaration as mandatory, its systematic absence reduces born-digital preservation conformance to zero across the entire corpus. This finding is the first demonstration of the indicator logic: a 0.0% conformance rate at this layer is not evidence of across-the-board documentary failure but of a specific, patchable encoding decision that is absent at scale. The measurement is interpretable precisely because the indicator translation pipeline connects the failure to its documentary function and locus of remediation, rather than leaving it as an undifferentiated zero.

At the document level, the cross-profile distribution is correspondingly narrow: 18 documents pass only PDF/A-1b, 10 pass only PDF/UA-1, and 1,312 pass none of the three profiles. The low overlap between the PDF/A-1b and PDF/UA-1 passing sets reinforces the two-level structure established in the theoretical framework: preservation viability and operational accessibility are distinct documentary properties that do not co-vary reliably in this corpus. The robustness assessment confirms the stability of this pattern. Excluding the three highest-contributing publishers reduces PDF/A-1b and PDF/A-2u compliance to 0.0%, while PDF/UA-1 compliance remains low at 1.1%. The layered nonconformance pattern, therefore, reflects corpus-wide infrastructure conditions rather than concentration effects from a small number of high-output programs.

4.2 Nonconformance is not homogeneous

The central methodological claim of the indicator framework is that nonconformance is not a single measurement state: different failures impair different documentary functions and imply different remediation logics. The debt-family activation data provide the empirical test of this claim. If nonconformance were homogeneous, all nine families would activate at similar rates across profiles. Instead, the observed pattern is highly differentiated—both across families within a profile and across profiles within a family—in ways that map systematically onto the documentary function each family represents.

Debt family activation is concentrated in two dominant categories. As shown in Table 3, Provenance & Metadata Debt activates in 97.5% of documents under PDF/A-1b, 100.0% under PDF/A-2u, and 97.8% under PDF/UA-1. Structural & Semantic Debt activates in 41.5% of documents under PDF/A-1b and

41.4% under PDF/A-2u but rises sharply to 98.4% under PDF/UA-1. Together, these two families account for the most persistent and pervasive forms of documentary debt in the corpus—yet they carry opposite interpretive implications. Provenance & Metadata Debt is post hoc patchable: the underlying obligation is a metadata declaration that can be injected into a finished file without restructuring the document. Structural & Semantic Debt under PDF/UA-1 is upstream-only: it reflects logical tag structure, reading order, and content-artifact classification decisions made—or not made—at the authoring stage, which cannot be reliably reconstructed from a completed PDF. A binary conformance indicator would report both families as "failed," obscuring this difference entirely.

Table 3
Debt family activation rates by conformance profile

Debt family	A-1b	A-2u	UA-1	Remediation locus
Provenance & Metadata	97.5%	100.0%	97.8%	Posthoc-patchable
Color & Output Profile	76.1%	75.7%	0.0%	Pipeline-fixable
File Integrity & Packaging	75.2%	6.3%	58.8%	Pipeline-fixable
Renderability & Resource	77.7%	44.6%	36.0%	Pipeline-fixable
Preservation Safety	0.1%	0.7%	0.0%	Pipeline-fixable
Structural & Semantic	41.5%	41.4%	98.4%	Upstream-only
Text Extractability & Unicode	47.3%	43.2%	43.2%	Pipeline-fixable
Language Discoverability	0.0%	0.0%	40.1%	Pipeline-fixable
Table Semantics	0.0%	0.0%	17.7%	Upstream-only

The profile-specific pattern for Provenance & Metadata Debt varies by triggering conditions across layers. Under PDF/A-2u, universal activation is driven by the missing PDF/A identification schema described in section 4. 4.1. Under PDF/UA-1, near-universal activation is driven primarily by missing or incomplete document title metadata required for accessibility-oriented interpretation. The debt family label is stable across profiles; the specific requirements that activate it are not. This stability-within-variation aligns with the family's construct definition—it measures a single documentary function (document identity and catalogability) through profile-specific rule sets—and illustrates why family-level, rather than rule-level, aggregation is the appropriate indicator unit.

The sharp profile effect in Structural & Semantic Debt reflects the qualitative expansion of requirements at the accessibility threshold. PDF/A profiles test a narrow set of structural properties relevant to self-containment; PDF/UA-1 adds requirements for logical tag structure, reading order, and content-artifact classification that are absent from preservation profiles. The median breadth for this family under PDF/UA-1 is four distinct rule violations per document, with a maximum of eight, indicating that

structural accessibility problems rarely involve a single missing element but typically span multiple interdependent encoding decisions.

Several other families clarify the distinct logic of the validation layers. File Integrity & Packaging Debt activates in 75.2% of documents under PDF/A-1b but only 6. 6.3% under PDF/A-2u—a reduction largely consistent with PDF/A-1b's prohibition of cross-reference streams, which PDF/A-2 explicitly permits. Color & Output Profile: Debt remains stable across the preservation profiles (76.1% and 75.7%) but disappears under PDF/UA-1, which does not test color-management requirements. These profile-specific disappearances and emergences are not anomalies; they confirm that each profile operationalizes a distinct documentary obligation and that the indicator system correctly tracks which families belong to which obligation.

Language Discoverability Debt and Table Semantics Debt are absent under both preservation profiles and emerge only under PDF/UA-1, where natural-language identification and semantic table structure become explicit requirements. Text Extractability & Unicode Debt is the only family that remains comparably stable across all three profiles (47.3%, 43.2%, and 43.2%), indicating that text-layer reliability is not specific to either preservation or accessibility obligations but cuts across both dimensions of documentary quality. Taken together, these patterns confirm the core claim of the indicator framework: nonconformance is not homogeneous, and the nine-family typology performs its intended function of differentiating documentary obligations that binary conformance rates conflate.

4.3 Disciplinary variation in latent documentary quality

If the debt-family typology functions as a valid multi-dimensional indicator system, different families should exhibit distinct patterns of disciplinary variation: families that measure infrastructure-wide structural properties should activate broadly across all disciplines, while families sensitive to field-specific encoding pressures should show selective differentiation. The disciplinary results are consistent with this expectation.

Under PDF/UA-1, Structural & Semantic Debt exceeds 94% in every disciplinary category in the corpus. This near-ceiling activation across all fields confirms that structural accessibility problems are not limited to specific disciplinary communities or production contexts; they reflect a corpus-wide infrastructural condition. The indicator does not discriminate between disciplines here because the underlying problem—the systematic absence of document-level semantic tagging—does not vary across them.

More discriminating patterns appear in Text Extractability & Unicode Debt and Language Discoverability Debt. Text Extractability & Unicode Debt shows the widest disciplinary range, varying from 7% in Agriculture to 74% in Science and Mathematics, with elevated rates also in Technology and Engineering (56%) and Language and Literature (63%), and substantially lower rates in Fine Arts (11%). These differences are consistent with the hypothesis that fields relying on specialized notation, multilingual

content, or non-standard character sets place greater pressure on encoding, font handling, and Unicode mapping—pressures that activate this family selectively rather than uniformly.

Language Discoverability Debt varies from 4% in Law to 74% in Fine Arts. Library and Information Science records the lowest rate among multi-journal categories (5%) for this family, despite reaching 100% activation for Structural & Semantic Debt. This contrast between near-zero Language Discoverability Debt and near-ceiling Structural & Semantic Debt within the same disciplinary category demonstrates that stronger performance on one documentary quality dimension does not entail stronger performance on another—a result consistent with the construct validity claim that the two debt vectors measure analytically independent properties.

Table Semantics Debt, which activates only under PDF/UA-1, is highest in Library and Information Science (44%) and Technology and Engineering (33%), consistent with the possibility that table-intensive document forms are more prevalent in these fields while semantic table encoding remains unevenly implemented. The disciplinary variation in this family is therefore plausibly content-driven rather than infrastructure-driven.

Taken together, the disciplinary results support the discriminant validity of the indicator system. Families capturing upstream structural properties exhibit near-uniform activation, indicating infrastructure-level problems that do not vary across fields. Families that capture encoding-sensitive or content-sensitive properties exhibit selective variation that tracks recognizable disciplinary differences in document production. This differentiated behavior is evidence that the typology measures distinct documentary dimensions rather than a single, undifferentiated quality signal.

4.4 Institutional variation and robustness to high-output publishers

The corpus includes 80 journals (71%) published by LPC member programs in the United States and 32 journals (29%) published by programs in Canada, corresponding to 926 and 414 documents, respectively. Across both national groups, conformance is very limited at all three validation layers, and the corpus-level pattern of near-universal noncompliance at the born-digital preservation floor and operational accessibility threshold is evident in both groups.

The clearest difference between the two groups appears at the legacy archival baseline. Canadian programs record a PDF/A-1b compliance rate of 3.9% (16 of 414 documents), whereas United States programs record a rate of 0.2% (2 of 926 documents). In the opposite direction, all 10 documents that pass PDF/UA-1 originate from United States programs, while Canadian programs record no PDF/UA-1 compliance. Neither group achieves any PDF/A-2u compliance. The two groups, therefore, differ in direction depending on the conformance layer considered—a result that reinforces the two-level structure of the indicator framework: national-group differences in preservation-oriented conformance do not predict differences in accessibility-oriented conformance.

At the debt-family level, the two groups are highly similar across the categories that dominate the corpus. Under PDF/UA- 1, Structural & Semantic Debt activates in 98.4% of documents from United States programs and 98.3% of documents from Canadian programs; Provenance & Metadata Debt activates in 98.3% and 96.6%, respectively. More visible, though still moderate, differences appear in Language Discoverability Debt (49.8% in Canada versus 35.9% in the United States) and Renderability & Resource Debt (43.7% versus 32.5%). The dominant debt families—those measuring infrastructure-wide properties—show no meaningful national differentiation. Differentiation is again confined to the selective families, consistent with the disciplinary patterns reported in section 4. 4.3.

The robustness assessment supports the stability of the overall pattern. Excluding the three highest-contributing publishers, which together account for 469 of the 1,340 documents, leaves the directional structure unchanged: PDF/A-1b and PDF/A-2u compliance fall to 0.0% in the reduced corpus. In contrast, PDF/UA-1 compliance remains low at 1.1%. The dominant debt families retain near-universal activation in the reduced corpus, confirming that the corpus-level findings reflect infrastructure-wide conditions rather than the output profile of a small number of high-volume programs. Full institution-level results are provided in the supplementary materials deposited in the OSF repository.

5 Discussions

The study's central methodological implication is that binary conformance indicators are insufficient. A failed PDF validation result becomes analytically meaningful only when mapped to the impaired documentary function and the implied remediation locus. The findings reported here demonstrate this principle across a corpus of 1,340 version-of-record PDFs: nonconformance is not a homogeneous state and treating it as such collapses analytically distinct problems into a single undifferentiated metric. The audit-based technical-debt framework developed here addresses this limitation by converting assertion-level validation output into a nine-family typology that distinguishes preservation-related and accessibility-related obligations and classifies each by remediation locus.

The framework demonstrated library-based open-access publishing because this setting combines strong normative commitments to access and preservation with operationally heterogeneous production workflows. This combination makes the gap between institutional intent and documentary output empirically visible. The corpus is not a representative sample of library publishing quality, nor is the study designed to render a verdict on publishing programs. It is a demonstration context for a scalable measurement framework: a setting in which the gap between what publishing infrastructure is positioned to guarantee and what the documentary record actually contains can be measured and interpreted.

The most consequential individual finding is the complete absence of PDF/A-2u conformance. At first glance, a 0.0% pass rate may appear to indicate a severe, generalized preservation failure. More precisely, however, it functions as a structural signal: in this corpus, the born-digital preservation floor is missed not because of many unrelated defects but because a single mandatory requirement—the PDF/A

identification schema declaration in the XMP metadata block—is systematically absent. This is relevant for indicator interpretation because the problem is patchable post hoc. It does not require redesign of author workflows or upstream changes in document production practices; it points to a missing pipeline-level intervention at the export or post-processing stage. Bedford et al.(2024) describe a closely related gap in their analysis of LPC preservation practices: preservation expertise within libraries frequently remains disconnected from the publishing workflows through which it could be applied. The 0.0% PDF/A-2u result is interpretable as an instance of that disconnection at the document level—a finding that is only legible because the indicator framework separates remediation locus from conformance status.

The accessibility findings are structurally more demanding. Under PDF/UA-1, Structural and Semantic Debt is present in 98.4% of documents, and the dominant deficiencies are precisely those classified as upstream-only: logical tag structure, reading order, and content–artifact distinction. Unlike the missing PDF/A identification schema, these problems cannot be reliably repaired after PDF generation. They depend on authoring-stage decisions, template design, and platform configurations that shape the document before export. This is the second major indicator distinction the framework makes visible: the two dominant debt families in the corpus—Provenance and Metadata Debt and Structural and Semantic Debt—are superficially similar in prevalence but carry opposite remediation implications. One can be corrected at scale through a pipeline intervention; the other requires upstream changes that touch the document before it is ever published. A system that reported only conformance rates would render this distinction invisible.

A boundary condition of the framework deserves explicit acknowledgment. Automated validation captures structurally detectable failures with precision and at scale, but it cannot substitute for user-centered accessibility evaluation (Cooper et al., 2012; Vollenwyder et al., 2023). The framework does not claim that assertion-level conformance exhausts the scope of documentary quality or lived accessibility. Rather, it treats automated validation output as a scalable infrastructural audit layer: a first-pass indicator system that identifies which categories of documentary obligation are systematically unmet, at a level of granularity that existing science metrics do not provide. Where validation results indicate upstream-only structural deficits, the implication is not that remediation is technically impossible, but that it requires different interventions—and different infrastructure-level decisions—than post hoc patchable failures.

The disciplinary and national comparisons are most useful for what they reveal about the indicator system's behavior rather than about disciplinary communities themselves. The near-ceiling activation of Structural and Semantic Debt across all disciplinary categories confirms that this family captures an infrastructure-wide property: the absence of semantic accessibility structures is not a feature of specific scholarly fields, but a condition distributed across the corpus regardless of subject domain. This is consistent with the family's upstream-only remediation locus—the problem arises at the platform and template level, shared across journals regardless of disciplinary content. By contrast, the wider disciplinary variation in Text Extractability, Unicode Debt, and Language Discoverability Debt indicates

that these families are more sensitive to field-specific encoding pressures. This differentiated behavior is a property of the indicator system, not merely of the publishing landscape: it demonstrates that the typology is measuring distinct documentary functions rather than a single composite quality signal.

The broader implication for science metrics follows directly from this differentiation. Existing indicator systems assess whether scholarly outputs are reachable, indexed, and cited; they do not assess whether those outputs remain archivally stable and machine operable as documentary artifacts. OA status confirms that a document is accessible; DOAJ indexing confirms editorial transparency; citation counts confirm uptake. None of these indicators is sensitive to whether the PDF carrying the scholarly record will be renderable across software generations, processable by text-mining systems, or navigable by assistive technologies. The framework developed here is not a replacement for those indicators; it operates at a different measurement level. But it demonstrates that the documentary quality of the version of record is a tractable empirical question—one that can be operationalized reproducibly, decomposed into interpretable dimensions, and connected to specific remediation logic. The patchable–upstream distinction is the analytical hinge on which this tractability depends: without it, the only interpretable output is a conformance rate, and the only available response is a general injunction to improve.

6 Limitations

Several features of the study design limit the scope of the findings. The corpus is restricted to LPC member programs in the United States and Canada with active DOAJ-indexed journals, representing a formalized, quality-oriented subset of library-based publishing; findings should not be assumed to generalize to unaffiliated, non-indexed, or non-Anglophone publishing contexts. The cross-sectional retrieval window (2024–2025) captures current production conditions but precludes assessment of longitudinal trends in conformance.

Three methodological constraints bear on the indicator system itself. First, documentary quality is operationalized through two observable dimensions—preservation viability and operational accessibility—assessed via PDF/A and PDF/UA conformance profiles; the construct may have additional dimensions not captured by conformance validation. Second, the nine-family debt typology is an interpretive analytical framework derived from automated rule-to-family mapping and manual review of edge cases, not a definitive classification system; the remediation-locus assignments are generalizations from the technical literature rather than empirically verified workflow mappings. Third, the intensity indicator is subject to a document-length confound—families that generate repeated assertions by design will produce higher values for longer documents regardless of problem severity—which is why activation serves as the primary analytical unit throughout.

Finally, automated conformance checking captures structurally detectable failures but cannot substitute for user-centered accessibility evaluation; the debt families reported here represent a lower bound on documentary quality problems rather than an exhaustive inventory (Cooper et al., 2012; Vollenwyder et

al., 2023). The study does not include workflow interviews, platform audits, or authoring-environment analysis; links between observed debt patterns and specific production decisions remain interpretive rather than causally demonstrated.

7 Conclusion

This study develops and demonstrates a reproducible audit-based framework for measuring latent documentary quality in scholarly publishing infrastructures. Evaluated against three successive conformance profiles—PDF/A-1b as a legacy archival baseline, PDF/A-2u as a born-digital preservation floor, and PDF/UA-1 as an operational accessibility threshold—a corpus of 1,340 version-of-record PDFs from DOAJ-indexed journals associated with Library Publishing Coalition member programs shows near-universal noncompliance at every layer. The methodological contribution of the study lies not in the conformance rates themselves, but in what the indicator framework reveals about the structure of that noncompliance.

The central finding is that nonconformance is not a homogeneous measurement state. Some PDF validation failures indicate patchable metadata omissions—in this corpus, the missing PDF/A identification schema declaration that reduces PDF/A-2u conformance to zero across all 1,340 documents. Others indicate upstream structural deficits—the near-universal absence of logical tag structure, reading order, and semantic accessibility encoding—which produce 98.4% Structural and Semantic Debt activation under PDF/UA-1. These two failure types look identical in a binary pass/fail indicator; they are analytically and practically distinct. The nine-family technical-debt typology operationalizes this distinction by classifying each debt family according to the impaired documentary function and the implied remediation locus, thereby converting machine-checkable violations into interpretable quality dimensions.

The measurement framework contributes to a gap in existing science metrics. Current indicators assess whether scholarly outputs are open, indexed, and cited. They do not assess whether the documentary object carrying the scholarly record is archivally stable, machine-operable, and accessible to the systems and users that depend on it. A version of the record can be openly available, DOAJ-indexed, DOI-assigned, and widely cited. At the same time, its PDF remains structurally opaque to assistive technologies, is recoverable only under legacy rendering conditions, and is unprocessable by text-mining pipelines. The audit-based indicator framework developed here addresses this blind spot by demonstrating that document-level quality is a tractable empirical property: it can be operationalized through conformance profiles, decomposed into functionally distinct debt families, and connected to specific points in the production workflow where intervention is feasible.

The framework is not restricted to library publishing. The indicator construction logic—translating machine-validation output into functionally interpreted quality dimensions, classified by documentary function and remediation locus—applies to any documentary corpus for which conformance profiles can be defined. National repositories, commercial publisher platforms, preprint servers, journal hosting

systems, and institutional repositories all produce version-of-record PDFs whose documentary quality has not been assessed by existing metrics. Future work may extend the framework to these contexts, enabling comparative assessment of infrastructure quality across publishing models and at scales not currently supported by existing science metrics. The present study offers a replicable baseline and a construction model for that extension.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics statement

This study did not involve human participants, animals, or personally identifiable data. Therefore, ethics approval was not required.

Author Contribution

Kemal Yayla: Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

References

1. Bedford, E., Dufour, C., Guimont, C., Howard, R., Nackerud, S.: Preserving what we publish: Findings from the library publishing coalition preservation task force. Coll. Res. Libr. News. **85**(1) (2024). https://doi.org/10.5860/crln.85.1.32
2. Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., Guðnason, G.: Open Access to the Scientific Journal Literature: Situation 2009. PLoS ONE. **5**(6), e11273 (2010). https://doi.org/10.1371/journal.pone.0011273
3. Buckland, M.K.: What Is a Document? J. Am. Soc. Inform. Sci. **48**(9), 804–809 (1997). https://doi.org/10.1002/(SICI)1097-4571)48:9%3C804::AID-ASI5%3E3.0.CO;2-V
4. Buckland, M.K.: Documentality Beyond Documents. Monist. **97**(2), 179–186 (2014). https://doi.org/10.5840/monist201497212
5. Buckland, M.K.: Document Theory. Knowl. Organ. **45**(5), 425–436 (2018). https://doi.org/10.5771/0943-7444-2018-5-425
6. Butler, S., Gamalielsson, J., Lundell, B., Brax, C., Mattsson, A., Gustavsson, T., Feist, J., Lönroth, E.: Maintaining interoperability in open source software: A case study of the apache PDFBox project. J. Syst. Softw. **159**, 110452 (2020). https://doi.org/10.1016/j.jss.2019.110452

7. Colman, J., Markovac, J.: Library journal publishing. In J. Markovac, M. Kleinman, and M. Englesbe (Eds.), *Medical and Scientific Publishing* (pp. 25–33). Elsevier. (2018). https://doi.org/10.1016/B978-0-12-809969-8.00004-8
8. Cooper, M., Sloan, D., Kelly, B., Lewthwaite, S.: A challenge to web accessibility metrics and guidelines: Putting people and processes first. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 1–4. (2012). https://doi.org/10.1145/2207016.2207028
9. Cunningham, W.: The WyCash Portfolio Management System. ACM SIGPLAN OOPS Messenger. **4**(2), 29–30 (1993). https://doi.org/10.1145/157710.157715
10. Darvishy, A.: PDF Accessibility: Tools and Challenges. In K. Miesenberger and G. Kouroupetroglou (Eds.), *Computers Helping People with Special Needs* (Vol. 10896, pp. 113–116). Springer International Publishing. (2018). https://doi.org/10.1007/978-3-319-94277-3_20
11. Darvishy, A., Sethe, R., Engler, I., Pierres, O., Manning, J.: The state of scientific PDF accessibility in repositories: A survey in Switzerland. Learn. Publish. (2023). https://doi.org/10.1002/leap.1581
12. Delgado-Quirós, L., Ortega, J.L.: Completeness degree of publication metadata in eight free-access scholarly databases. Quant. Sci. Stud. **5**(1), 31–49 (2024). https://doi.org/10.1162/qss_a_00286
13. Deville, J., Sondervan, J., Stone, G., Wennström, S.: Rebels with a cause? Supporting library and academic-led open access publishing. LIBER Quarterly: J. Association Eur. Res. Libr. **29**(1), 1–28 (2019). https://doi.org/10.18352/lq.10277
14. Drümmer, O., Erle, M.: PDF/UA – a New Era for Document Accessibility: Understanding, Managing and Implementing the ISO Standard PDF/UA (universal Accessibility): Introduction to the Special Thematic Session. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) Computers Helping People with Special Needs, vol. 7382, pp. 585–586. Springer, Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-31522-0_88
15. Dryden, J.: PDF/A-1: A Ray of Light in the Digital Dark Age? J. Arch. Organ. **6**(1–2), 121–124 (2008). https://doi.org/10.1080/15332740802246841
16. Erle, M., Hofer, S.: Validity and semantics – two essential parts of a backbone for an automated PDF/UA compliance check for PDF documents. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) Computers Helping People with Special Needs, vol. 7382, pp. 617–620. Springer, Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-31522-0_93
17. Fanning, B.: *Preservation with PDF/A (2nd Edition)*. Digital Preservation Coalition. (2017). https://doi.org/10.7207/twr17-01
18. Fischer, T., Lundell, B., Gamalielsson, J.: Achieving conformance to document standards: Can PDF files conform to the PDF/a-1b specification? Int. J. Stand. Res. **19**(1), 1–32 (2021). https://doi.org/10.4018/ijsr.288523
19. Fischer, T., Lundell, B., Gamalielsson, J.: On PDF/a conformance and font usage in PDF documents provided by public sector organizations. Int. J. Stand. Res. **20**(1), 1–19 (2023). https://doi.org/10.4018/IJSR.329605

20. Hinchliffe, L.J.: The version of record as a central organizing concept in scholarly publishing. Inform. Serv. Use. **42**(3–4), 309–314 (2022). https://doi.org/10.3233/ISU-220164
21. Kruchten, P., Nord, R.L., Ozkaya, I.: Technical debt: From metaphor to theory and practice. IEEE Softw. **29**(6), 18–21 (2012). https://doi.org/10.1109/MS.2012.167
22. Kumar, A., Wang, L.L.: Uncovering the new accessibility crisis in scholarly PDFs: Publishing model and platform changes contribute to declining scholarly document accessibility in the last decade. *The 26th International ACM SIGACCESS Conference on Computers and Accessibility*, 1–16. (2024). https://doi.org/10.1145/3663548.3675634
23. Lehtonen, J., Koivunen, K., Helin, H., Kylander, J.: *PDF mayhem: Is broken really broken?* iPRES 2018. (2018). https://doi.org/10.17605/OSF.IO/FZXC9
24. Levy, D.M.: Fixed or fluid? Document stability and new media. *Proceedings of the 1994 ACM European Conference on Hypermedia Technology - ECHT '94*, 24–31. (1994). https://doi.org/10.1145/192757.192760
25. Lippincott, S.K.: Library as publisher: New models of scholarly communication for a new era. Against the Grain (2017). https://doi.org/10.3998/mpub.9944345 LLC
26. Mattson, M., Friend, L.: A planning perspective for library journal publishing services. OCLC Syst. Services: Int. Digit. Libr. Perspect. **30**(3), 178–191 (2014). https://doi.org/10.1108/OCLC-01-2014-0005
27. McGuinness, R., Wilson, C., Johnson, D., Doubrov, B.: veraPDF: Open Source PDF/A Validation Through Pragmatic Partnership. *iPRES2017*. iPRES. (2017). https://ipres-conference.org/ipres17/ipres2017.jp/wp-content/uploads/28Rebecca-McGuinness.pdf
28. Miesenberger, K., Kobayashi, M., Peňáz, P. PDF accessibility in international academic publishers. In, Pierrès, O., Schmitt-Koopmann, F., and (eds.): *Lecture Notes in Computer Science* (pp. 38–46). Springer Nature Switzerland. (2024). https://doi.org/10.1007/978-3-031-62846-7_5
29. Nganji, J.T.: An assessment of the accessibility of PDF versions of selected journal articles published in a WCAG 2.0 era (2014–2018). Learn. Publish. **31**(4), 391–401 (2018). https://doi.org/10.1002/leap.1197
30. Piwowar, H., Priem, J., Larivière, V., Alperin, J.P., Matthias, L., Norlander, B., Farley, A., West, J., Haustein, S.: The state of OA: A large-scale analysis of the prevalence and impact of open access articles. PeerJ. **6**, e4375 (2018). https://doi.org/10.7717/peerj.4375
31. Pradhan, D., Rajput, T., Rajkumar, A.J., Lazar, J., Jain, R., Morariu, V.I., Manjunatha, V.: Development and evaluation of a tool for assisting content creators in making PDF files more accessible. ACM Trans. Accessible Comput. **15**(1), 1–52 (2022). https://doi.org/10.1145/3507661
32. Rajkumar, A.J., Lazar, J., Jordan, J.B., Darvishy, A., Hutter, H.-P.: PDF accessibility of research papers: What tools are needed for assessment and remediation? *53rd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. (2020). https://doi.org/10.24251/HICSS.2020.512

33. Rømen, D., Svanæs, D.: Validating WCAG Versions 1.0 and 2.0 Through Usability Testing with Disabled Users. Univ. Access Inf. Soc. **11**(4), 375–385 (2012). https://doi.org/10.1007/s10209-011-0259-3
34. Sandy, H.M., Mattern, J.B.: Academic library-based publishing: A state of the evolving art. Libr. Trends. **67**(2), 337–357 (2018). https://doi.org/10.1353/lib.2018.0040
35. Shanahan, D.R.: A living document: Reincarnating the research article. Trials. **16**(1), 151 (2015). https://doi.org/10.1186/s13063-015-0666-5
36. Skinner, K.E., Lippincott, S., Speer, J., Walters, T.: Library-as-publisher: Capacity building for the library publishing subfield. J. Electron. Publishing. **17**(2) (2014). https://doi.org/10.3998/3336451.0017.207
37. Smiraglia, R.P.: The nature of a work, 1st edn. Scarecrow (2001). (Original work published Bloomsbury Publishing)
38. Stapleton, S.C.: A team approach: Library publishing partnerships with scholarly societies. J. Librariansh. Sch. Communication. **7**(1) (2019). https://doi.org/10.7710/2162-3309.2326
39. Szentirmai, A.B., Inal, Y., Torkildsby, A.B.: The accessibility paradox: Can research articles inspecting accessibility be inaccessible? In K. Miesenberger, P. Peňáz, and M. Kobayashi (Eds.), *Computers Helping People with Special Needs* (Vol. 14750, pp. 47–54). Springer Nature Switzerland. (2024). https://doi.org/10.1007/978-3-031-62846-7_6
40. Tabarés, R.: Open access, responsibility and the platformization of academic publishing. NOvation - Crit. Stud. Innov. 2147 (2022). https://doi.org/10.5380/nocsi.v0i2.91157
41. Vollenwyder, B., Petralito, S., Iten, G.H., Brühlmann, F., Opwis, K., Mekler, E.D.: How compliance with web accessibility standards shapes the experiences of users with and without disabilities. Int. J. Hum. Comput. Stud. **170**, 102956 (2023). https://doi.org/10.1016/j.ijhcs.2022.102956
42. Wilson, C., McGuinness, R., Jung, J.: veraPDF: Building an open source, industry supported PDF/A validator for cultural heritage institutions. Digit. Libr. Perspect. **33**(2), 156–165 (2017). https://doi.org/10.1108/DLP-08-2016-0031