Advances in Laboratory Automation
Vol. 1 No. 1 (2026)

A preservation range can hide a year reported as in progress

Camille Taylor · Charles Carr · Gabriel Wright

Published: November 26, 2026 | Article 1211

Abstract

A preservation program can be named in an article record, listed for its journal, or reported as holding a particular issue. We compared these levels for 11,904 Crossref journal articles and public ISSN Keepers records from six serials and three publication years. Crossref named a provider on 7,732 articles. All 4,172 articles with no declaration had a serial-level Preserved span covering their year. Of 10,358 declared article–provider pairs, 8,252 matched a Preserved span and 1,807 were supported by an extent naming the year explicitly. The difference mostly came from extent descriptions that did not identify the year directly. One concrete gap was visible. For PeerJ Computer Science, the CLOCKSS Preserved span of 2015–2019 overlapped 2018, but the extent omitted 2018 and another row listed it as In Progress. Serial holdings and article declarations can therefore be reconciled more precisely when the reported status and extent survive the join.

A preservation range can hide a year reported as in progress

Abstract

A preservation program can be named in an article record, listed for its journal, or reported as holding a particular issue. We compared these levels for 11,904 Crossref journal articles and public ISSN Keepers records from six serials and three publication years. Crossref named a provider on 7,732 articles. All 4,172 articles with no declaration had a serial-level Preserved span covering their year. Of 10,358 declared article–provider pairs, 8,252 matched a Preserved span and 1,807 were supported by an extent naming the year explicitly. The difference mostly came from extent descriptions that did not identify the year directly. One concrete gap was visible. For PeerJ Computer Science, the CLOCKSS Preserved span of 2015–2019 overlapped 2018, but the extent omitted 2018 and another row listed it as In Progress. Serial holdings and article declarations can therefore be reconciled more precisely when the reported status and extent survive the join.

Keywords Digital preservation, serial holdings, ISSN Keepers Registry, Crossref, archival metadata

Introduction

Imagine a library staff member reviewing the preservation records of a journal before a platform change. One DOI record names an archive, another has no preservation field, and the serial's public holdings page puts the publication year inside a range classified as preserved. Those observations answer questions at different levels. The DOI field specifies where the publishing record declares an arrangement. The serial record says what an archive has reported about a continuing publication. Its first and last years can summarize a span containing several differently described components. For a volume or issue audit, the difference between a span and an explicit extent matters. An archive clause that reports a range can reach a different conclusion from the volume listing printed beside it.

The Digital Preservation Coalition's handbook positions preservation as a sustained set of activities (Digital Preservation Coalition, 2015). Hedstrom (1997) identified how electronic records depend on ongoing institutional and technical work. An early survey by Kenney et al. (2006) found that e-journal archiving arrangements varied in scope and that enrollment in individual services did not cover all parts of the serial landscape. Rieger (2018) describes continued gaps between preservation commitments and what organizations can verify over time. More recently, Laakso et al. (2021) traced vanished open-access journals to understand the practical consequences of discontinuity. Eve (2024) estimated article-level archival coverage using holdings data from multiple archives. Our study asks a narrower question about the information one can obtain from public preservation metadata before attempting item-level verification.

Crossref makes member-provided information available for most kinds of scholarly records (Hendricks et al., 2020). Its deposit schema allows an article record to identify archive organizations (Crossref, 2026). The agency reports shown in the ISSN Portal have a different source. The Keepers Registry collects reports about serial holdings and exposes them in the record for a serial's ISSN (ISSN International Centre, n.d.-a). Burnhill (2013) explains why tracking e-journal holdings at the serial level can help libraries monitor fragmented arrangements, and Tavernier et al. (2021) describes how the registry displays what institutional users can find about a title's coverage. An exact journal identifier can connect the records, but it cannot force an article-level statement and a serial-level report to represent the same property.

The record for Yi et al. (2024) illustrates how the fields can differ. The corresponding Crossref work has no archive names in its returned metadata, even though the public Keepers record for electronic ISSN 2059-3635 has four agencies with Preserved spans covering 2024. Only Portico offers a dated extent in the selected fields that expressly matches this work's recorded volume 9 and issue 1. Another archive's display uses volume ranges without dating each listed volume. The metadata for the article, the serial-year interval, and the dated issue are thus three administrative observations. None requires reading the publication or retrieving an archival copy.

The study asks how the number of identified archives changes as one requires a matching serial, a preserved temporal span, an explicitly named year, or a dated volume and issue. We also compare archive names declared in DOI records with those found in the serial record. Those questions distinguish a broad inventory useful for identifying agencies to ask from a narrower evidence check that a reported extent names the period in question.

Data and methods

Publication rosters and public holdings

Six deliberately selected online serials were queried for three full publication years, 2018, 2021, and 2024. The selection included the benchmark serial (2059-3635), Scientific Data (2052-4463), PeerJ Computer Science (2376-5992), the Journal of Open Source Software (2475-9066), eLife (2050-084X), and the Journal of Open Research Software (2049-9647). These journals supply repeated examples of differing registry configurations and publishers. The sample is not representative of all indexed serials. The intervening years are outside the fixed design. From the Crossref journal-works API we requested records with type journal-article and publication dates within the full calendar year, retaining only DOI, archive names, volume, and issue. The documentation states the distinction between the publicly queried journal and work endpoints (Crossref, 2020). We followed the journal cursor to the advertised result total for each journal-year and checked that the DOIs used in the annual panel were unique. A shorter December-only query for the same years provides a sensitivity check.

The preservation observations came from the ISSN Portal's public publisher pages, using the same online ISSN in each case. These pages expose serial-level statements as linked data, from which we retained archive component name and identifier, reported status, inclusive start and end years, extent text, and the record's agency update date. The component descriptions contain administrative extent tokens such as volume and issue numbers. No archived articles, attachments, or links to third-party files were followed. The Keepers participant documentation explains that individual agencies contribute holding metadata (ISSN International Centre, n.d.-b). Agency dates label their reports. They do not indicate the day a specific DOI was ingested. The findings refer to one later snapshot viewed on 23 November 2026 UTC, with the 2018, 2021 and 2024 labels referring to publication years.

An exact match on the electronic ISSN avoids guessing based on similar journal titles. The comparison also uses the same agency identifiers across repeated reports for one title. We retain a documented crosswalk for the named agencies that actually enter the Crossref declarations. For example, the schema name CLOCKSS and the Keepers agency key for CLOCKSS Archive identify the same participating program. Christen (2012) distinguishes matching by known identifiers from probabilistic resolution, and Van de Sompel et al. (2014) describe why persistent records must keep the referents of multiple identifiers clear. Here the DOI fixes an article, the eISSN fixes its online serial, and the archive key fixes a reporting agency. The join lets these entities be compared without substituting one unit for another.

Rules for comparing preservation statements

The first rule counts agencies listed for the serial at any time in the public record. An archive can enter that list even if its preserved coverage has no range containing the article's year. The second rule requires a component with status Preserved whose start and end years enclose the selected publication year. The range endpoints are inclusive. A separate indicator stores an In Progress interval rather than treating it as a preserved interval, and an agency is counted once for each serial and year even if several preserved components meet the rule. This avoids counting repeated reports from one agency or two publisher labels as separate archives. Reported status and date bounds alone form a serial-year condition, with no claim that the entire interior is enumerated.

The third rule looks for a year actually named in a Preserved extent. It accepts obvious four-digit year tokens in simple year lists, spans made explicit in the extent, and a dated volume such as 2024 followed by a colon and a numeric volume. It also handles commas inside parentheses so that issue lists do not start a new year entry. It does not infer years from bare volume numbers or assume that a title has a regular volume schedule. When every component with a preserved span covering the year supplies a fully interpretable year list and none names the year, we record a gap in the enumerated years. A bare volume list, a mixed extent that does not explicitly include the selected year, and a missing extent remain unresolved unless another component supplies the needed year. The ambiguity of converting between metadata representations is a recognized problem in interoperability (Haslhofer & Klas, 2010). Whether a description is of sufficient quality depends on what a user needs to recover (Park, 2009).

The fourth and fifth rules look within dated extents for the exact numeric volume and then, when present, the exact numeric issue of the Crossref work. A clear range of issues can qualify if it contains the observed issue; a volume with no issue enumeration cannot satisfy that check. These locator rules do not apply to a list of undated journal volumes, although a human using more contextual information might be able to interpret such a list. The KBART standard provides context for title-list and holdings data at levels below the title (National Information Standards Organization, 2026). van Ballegooie et al. (2017) discuss the role of exact online issue and volume information in coverage descriptions. Our analysis compares the publicly exposed extent strings and does not require that Keepers JSON-LD or Crossref use the KBART tabular format.

We count both article records and agency-by-year reports. In an article–agency pair the same article contributes one row for each provider named in the Crossref archive field. This outcome reflects how many declarations would change under each rule in the selected work roster. At the serial level an agency contributes one distinct entry in each of the three years regardless of the number of DOIs issued. This outcome exposes the scope of the holdings information separately from publisher volume. The comparison reports paired descriptive counts without treating repeated article declarations within a journal as independent evidence.

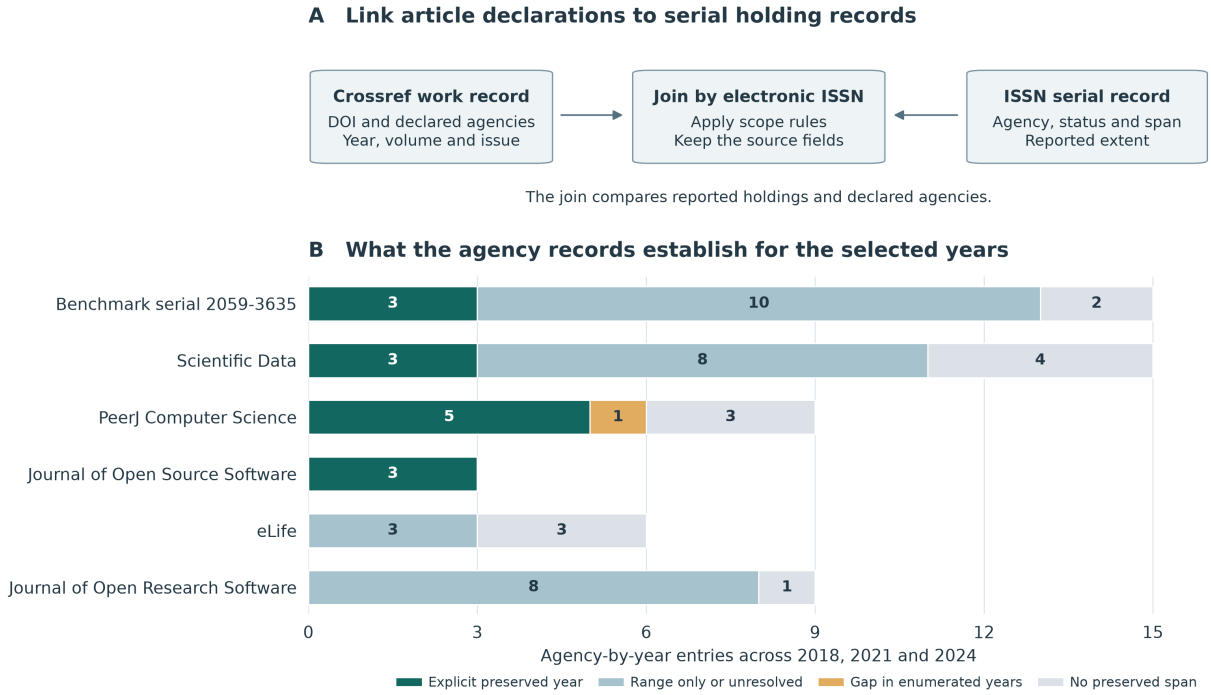

Figure 1. The model connects article declarations and serial holdings through the matching electronic ISSN. The lower panel partitions agency–year entries by whether the preserved report expressly names the year, leaves it unresolved within a span, omits it from a fully readable enumeration, or supplies no preserved span.

Results

The annual collection contained 11,904 unique Crossref articles. A deposit named one or more archive agencies on 7,732 works. These came from eLife, which named CLOCKSS on all 6,419 works, and PeerJ Computer Science, which named CLOCKSS, LOCKSS and Portico on all 1,313 works. The remaining 4,172 works, from four serials, did not provide a declaration. Every work had at least one matching agency-reported Preserved span at its serial and year, including every work with no Crossref archive field. That is a consistent property of the six deliberately selected serials and years rather than a coverage estimate for the registry.

Table 1. Article declarations and positive registry evidence by serial

Serial	Articles	Archive declaration	At least one Preserved span	Explicit preserved year	Dated year and volume	Dated year, volume and issue
Benchmark serial	845	0	845	845	845	845
Scientific Data	2,077	0	2,077	2,077	2,077	2,077
PeerJ Computer Science	1,313	1,313	1,313	1,313	1,308	0
Journal of Open Source Software	1,173	0	1,173	1,173	1,173	1,173
eLife	6,419	6,419	6,419	0	0	0
Journal of Open Research Software	77	0	77	0	0	0
Total	**11,904**	**7,732**	**11,904**	**5,408**	**5,403**	**4,095**

Note. Each registry column counts articles with at least one agency passing the corresponding rule. A zero in an explicit-locator column does not resolve whether any relevant archive holds that item. The numeric issue field could not be used on 7,747 articles.

The date-specific results differed markedly from the span-based figures in the same roster. At least one agency named the year explicitly on 5,408 articles. In 5,403 of those, a date-qualified volume matched the Crossref volume. The difference came from five PeerJ Computer Science records. An exact numeric issue matched on 4,095 records. The date-qualified source for PeerJ Computer Science named volumes without supplying issues in the enrolled extent. For eLife and the Journal of Open Research Software, the relevant years remained unresolved under the explicit-year rule. Their reports contain archive spans and volume-style extent information without naming the three analyzed years in a form usable for the narrower comparison.

Table 2. Registry evidence for agencies declared in the matched Crossref article

Rule or observation	Annual work–agency pairs	December work–agency pairs
Declared in the Crossref archive field	10,358	968
Listed for the serial anywhere in Keepers	10,358	968
Covered by a Preserved span	8,252	784
Year explicitly named in a Preserved extent	1,807	137
Enumerated year gap inside a Preserved span	26	0
Unresolved year within a Preserved span	6,419	647

Note. The last three rows partition the Preserved-span row. A work contributes a separate pair for each declared agency. The December cohort is a subset of the corresponding annual work records.

Across the 18 serial-year frames, there were 57 possible agency-year entries using the agencies listed anywhere for each serial. A Preserved span covered 44 of them. Fourteen included an explicit preserved year, 29 remained unresolved from the reported extent, and one had a visible gap in a fully enumerated list. This distinction explains much of the difference between the article-weighted rules. The unresolved

eLife CLOCKSS account contributes 6,419 declaration pairs to the Preserved-span row while supplying no explicit year evidence for those three periods. The shorter December frames produce the same qualitative distinction, even though no PeerJ Computer Science articles were returned for December 2018 and the consequent 26 article–agency gap cases occur only in the annual panel.

The enumerated gap occurs in PeerJ Computer Science. The CLOCKSS component labeled Preserved provides endpoints 2015 and 2019, but its extent explicitly names 2015–2017 and 2019. A separate component labels 2018 In Progress and names it in its extent. All 26 Crossref articles in that serial's 2018 frame declare CLOCKSS among their providers. The span rule counts those names as covered. The explicit-year rule leaves those 26 outside the preserved enumeration. In the 2024 PeerJ Computer Science frame, all 819 records still name the three agencies, whereas Portico is the only one with a Preserved span covering 2024 in the captured registry. The comparison therefore distinguishes a year absent from an enumerated preserved row from a year beyond the recorded preserved endpoint.

Discussion

The agency name in a DOI deposit can direct further investigation. It becomes a different observation when joined to a serial-year span and still another when backed by a dated volume or issue in the registry. Bequet (2022) tracks how e-journal holdings are communicated across publisher and archiving systems. Cramer et al. (2023) discuss arrangements needed to preserve electronic scholarship over time. Their discussion provides context for the observed mismatch between work-level and serial-level sources. A blank field on an article record can mean that an optional deposit element was omitted. Production choices can affect how frequently optional metadata appear in Crossref (de Jonge & Kramer, 2026). In the selected rosters, the 4,172 blank fields did not prevent the serials from reporting current Preserved spans. This matters when an automated audit uses the work field as its only trigger for checking the public holdings record.

The PeerJ example suggests how a compact range can change the meaning of a record during export. A direct serial join produces a quick agency list. If next it assigns all years between two endpoints, it makes 2018 look like other years in the CLOCKSS component. The detail list gives a more discriminating account for this year. The printed interval is still useful as the bounds of a reported set. A machine-readable export can preserve both the interval and the explicit extent so that a later system can tell them apart. CLOCKSS provides a public holdings report and exchanges data with Keepers (CLOCKSS, n.d.). Portico itself distinguishes counts of committed titles from preserved titles and individual archived units (Portico, n.d.). Neither arrangement is reduced to the metadata comparison performed here.

Publication preservation involves copies, checks, and organizational arrangements that have a different evidentiary status from holding records. Rosenthal et al. (2005) identify system requirements for retaining digital objects over time. Maniatis et al. (2003) describe a method for assessing and repairing independently held replicas, which operates on preserved materials themselves. Keller (2020) reviews roles institutions and archival services can play. These perspectives support a clear boundary between checking a register's reported extent and checking preservation in action. A dated issue match is enough to trace a holdings statement, but it cannot tell whether a particular deposited article, its supplementary files, or an author's corrections were ingested and remain recoverable. A study on URI persistence by Jones et al. (2016) shows another difference between the continuing presence of an identifier and the behavior of the addressed resource. The present study measures record statements without requesting the underlying items.

Scope also changes with the counting unit. The 6,419 unresolved eLife links are many article declarations repeating one agency configuration over three title-years. Their weight is appropriate for describing the output of an article-by-article audit on this corpus. An assessment of how frequently different serials report year-qualified extents has three separate eLife agency-year observations in this design. It should not treat those thousands of work rows as independent evidence. Multi-index studies such as Visser et al. (2021) show that even broad public sources differ in record coverage. Our six-title panel can reveal where that distinction matters without suggesting an industry-wide proportion.

Lastly, the method is conservative about interpreting an extent. An unparsed volume-only listing is genuinely different from an enumerated year gap. Alternative data can sometimes date volumes or enable more granular checks, but inferring a calendar mapping from a regular publishing schedule would add assumptions to the output. The analysis also uses online ISSNs only, so reported coverage of another media version is outside its joins. It cannot determine how recently a particular item was ingested or what access terms apply to a library. The archived results give a fixed way to reproduce the differences in what the two reporting layers say. A follow-up could compare successive registry snapshots against directly supplied agency inventories and isolate changes in dates, status, or extent wording.

Conclusion

Six serials with positive Preserved spans in all three selected years still had substantial differences between declared providers, years named inside extents, and dated issue matches. The contrast is largest when a broad title-level report is repeated for many articles. The administrative record also supplies a specific 2018 discrepancy where a year inside a preserved range is explicitly placed in progress elsewhere in the same serial record. Keeping the source status and extent alongside the Crossref declaration preserves those differences for later verification.

Data and code availability

The companion archive retains the Crossref journal projections, the serial archive component fields, the exact benchmark identity, and source checksums. The scripts reproduce the comparison and Figure 1 in PNG and SVG. Author, funding, and competing-interest declarations remain for author confirmation.

References

Digital Preservation Coalition (2015). *Digital Preservation Handbook (2nd ed.)*. https://www.dpconline.org/handbook

Laakso, M., Matthias, L., & Jahn, N. (2021). Open is not forever: A study of vanished open access journals. *Journal of the Association for Information Science and Technology, 72*(9), 1099–1112. https://doi.org/10.1002/asi.24460

Jones, S. M., Van de Sompel, H., Shankar, H., Klein, M., Tobin, R., & Grover, C. (2016). Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content. *PLOS ONE, 11*(12), e0167475. https://doi.org/10.1371/journal.pone.0167475

Rosenthal, D. S. H., Robertson, T., Lipkis, T., Reich, V., & Morabito, S. (2005). Requirements for Digital Preservation Systems. *D-Lib Magazine, 11*(11). https://doi.org/10.1045/november2005-rosenthal

National Information Standards Organization (2026). *KBART: Knowledge Bases and Related Tools Recommended Practice (NISO RP-9-2026)*. https://doi.org/10.3789/niso-rp-9-2026

Park, J. R. (2009). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly, 47*(3-4), 213–228. https://doi.org/10.1080/01639370902737240

Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer. https://doi.org/10.1007/978-3-642-31164-2

Yi, L., Pan, H., Ning, Z., Xu, L., Zhang, H., Peng, L., Liu, Y., Yang, Y., Si, W., Wang, Y., Zhu, X., Huang, S., Meng, Z., & Xie, J. (2024). Clinical and biomarker analyses of SHR-1701 combined with famitinib in patients with previously treated advanced biliary tract cancer or pancreatic ductal adenocarcinoma: a phase II trial. *Signal Transduction and Targeted Therapy, 9*(1), 347. https://doi.org/10.1038/s41392-024-02052-3

Crossref (2026). *Archive locations*. https://www.crossref.org/documentation/schema-library/markup-guide-metadata-segments/archive-locations/

de Jonge, H. & Kramer, B. (2026). Manuscript submission systems and metadata completeness in Crossref: Patterns and associations. *PLOS ONE, 21*(3), e0345417. https://doi.org/10.1371/journal.pone.0345417

Van de Sompel, H., Sanderson, R., Shankar, H., & Klein, M. (2014). Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping. *International Journal of Digital Curation, 9*(1), 331–342. https://doi.org/10.2218/ijdc.v9i1.320

Kenney, A. R., Entlich, R., Hirtle, P. B., McGovern, N. Y., & Buckley, E. L. (2006). *E-journal archiving metes and bounds: A survey of the landscape*. Council on Library and Information Resources. https://www.clir.org/pubs/reports/pub138/

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies, 2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Keller, S. (2020). Demystifying Digital Preservation: Recommendations for Organizations, Libraries, and Information Professionals. *The Serials Librarian, 78*(1-4), 57–63. https://doi.org/10.1080/0361526x.2020.1697591

Bequet, G. (2022). From the Cradle to the Digital Vault: Tracking the Path of E-journals. *The Serials Librarian, 82*(1-4), 199–204. https://doi.org/10.1080/0361526x.2022.2028498

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies, 1*(1), 414–427. https://doi.org/10.1162/qss_a_00022

Burnhill, P. (2013). Tales from The Keepers Registry: Serial Issues About Archiving & the Web. *Serials Review, 39*(1), 3–20. https://doi.org/10.1080/00987913.2013.10765481

Crossref (2020). *REST API*. https://www.crossref.org/documentation/retrieve-metadata/rest-api

Rieger, O. Y. (2018). *The state of digital preservation in 2018: A snapshot of challenges and gaps*. Ithaka S+R. https://doi.org/10.18665/sr.310626

Maniatis, P., Rosenthal, D. S. H., Roussopoulos, M., Baker, M., Giuli, T., & Muliadi, Y. (2003). Preserving peer replicas by rate-limited sampled voting. *Proceedings of the nineteenth ACM symposium on Operating systems principles*, 44–59. https://doi.org/10.1145/945445.945451

Portico (n.d.). *Facts and figures*. https://www.portico.org/coverage/facts-and-figures/

CLOCKSS (n.d.). *Archived content*. https://clockss.org/whats-in-clockss/archived-content/

ISSN International Centre (n.d.-a). *Keepers Registry*. https://www.issn.org/services/online-services/keepers-registry/

Hedstrom, M. (1997). Digital Preservation: A Time Bomb for Digital Libraries. *Computers and the Humanities, 31*(3), 189–202. https://doi.org/10.1023/a:1000676723815

Eve, M. P. (2024). Digital Scholarly Journals Are Poorly Preserved: A Study of 7 Million Articles. *Journal of Librarianship and Scholarly Communication, 12*(1). https://doi.org/10.31274/jlsc.16288

Tavernier, W., Westervelt, T., & Carlson, A. J. (2021). Where Do We Keep That? The New Keepers Registry and the Digital Content in Your Collection. *The Serials Librarian, 80*(1-4), 155–160. https://doi.org/10.1080/0361526x.2021.1865020

van Ballegooie, M., Meares, S., & Wilson, K. (2017). Deep Dive into KBART. *The Serials Librarian, 72*(1-4), 15–25. https://doi.org/10.1080/0361526x.2017.1309826

ISSN International Centre (n.d.-b). *The keepers*. https://keepers.issn.org/keepers

Cramer, T., German, C., Jefferies, N., & Wise, A. (2023). A perpetual motion machine: The preserved digital scholarly record. *Learned Publishing, 36*(2), 312–318. https://doi.org/10.1002/leap.1494

Haslhofer, B. & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys, 42*(2), 1–37. https://doi.org/10.1145/1667062.1667064