Annals of Adaptive Measurement
Vol. 1 No. 1 (2026)

A Recent Index Timestamp Does Not Identify a Recent Metadata Deposit

Eric Robertson · George Barrett · Noah Spencer

Published: November 26, 2026 | Article 1204

Abstract

A synchronization query for February 2024 publications returned 111 records when the date restriction used the Crossref index timestamp and none when it used the deposit timestamp. Every record in the 334-DOI cohort was eligible under the same publication frame. The date restriction began on 1 October 2026 and was checked both locally and against the public API. At an earlier cutoff, 145 recent deposits and 322 recent index entries were observed. Most of the recent index entries included material whose last visible deposit was outside the same window. The discrepancy is a property of the timestamps as exposed in this audit. It affects how the size and meaning of an incremental update batch can be reported.

A Recent Index Timestamp Does Not Identify a Recent Metadata Deposit

Abstract

A synchronization query for February 2024 publications returned 111 records when the date restriction used the Crossref index timestamp and none when it used the deposit timestamp. Every record in the 334-DOI cohort was eligible under the same publication frame. The date restriction began on 1 October 2026 and was checked both locally and against the public API. At an earlier cutoff, 145 recent deposits and 322 recent index entries were observed. Most of the recent index entries included material whose last visible deposit was outside the same window. The discrepancy is a property of the timestamps as exposed in this audit. It affects how the size and meaning of an incremental update batch can be reported.

Which event is recent?

Downloading records that changed recently sounds like one operation until the metadata service offers several clocks. An initial registration identifies the first appearance of a DOI in the service. A later deposit can replace or augment its registered description. A subsequent indexing event can reflect processing or a source-side change from somewhere else. A cutoff matched to one of those clocks can make a very old publication appear in the administrative update queue. Another cutoff with the same date can make the record disappear from that queue. The question is how to determine what such a count can say about the maintenance activity observed on a fixed set of journal publications.

Crossref distinguishes synchronization filters for new registrations, current deposits, and changes reflected in the service's index (Crossref, n.d.-b) (Crossref, n.d.-a). Its description of the metadata infrastructure emphasizes that different systems deposit and use the field contents (Hendricks et al., 2020). Studies of journal metadata practices show why publisher-level and system-level variation matter when looking at administrative information at article scale (Jahn et al., 2021) (de Jonge and Kramer, 2026). There is also a broader infrastructure question. Synchronization standards describe workflows in which a consumer updates a local copy based on representation, event time, or a published change list (Klein et al., 2013) (Open Archives Initiative, 2017). Those signals serve distinct needs depending on what an upstream service exposes. The traffic generated by refreshes can differ from the evidence for fresh contributor submissions.

We take a fixed-publication-month cohort and compare event-window selection for each of the three date fields in the same Crossref records. Then we check whether the local filter based on the projected dates reproduces independently submitted public API queries at the chosen synchronization cutoff. This design makes each outcome a property of the same DOI set. It isolates what changes when a user alters the definition of recency. It cannot determine what caused an index event or the number of intervening updates to each record. In particular it measures a separation between observable latest timestamps, with no claim that the index took that time to ingest a deposit.

An audit of the three public clocks

We retrieved the full Crossref publication-date results from 1 to 29 February 2024 for ESMO Open, PeerJ Computer Science, Scientific Data, and Machine Learning: Science and Technology. The journals were selected to provide examples across several publishing contexts, including the one containing the administrative example discussed below. The broad journal query uses each registered ISSN. The month field defines the population. The target outcome instead comes from the record's created, deposited, and indexed metadata. The acquisition retained only these dates and DOI, so no title, content, author information, or downstream performance indicator enters the analytical data.

Two screens keep the comparison interpretable. The first validates a well-formed single DOI in the result, with no duplicates within or between the four frame queries. The second preserves all four fully enumerated journal cohorts, regardless of which dates pass the subsequent filters. An exact object-level key fixes the records under comparison. Traditional record linkage and contemporary preprocessing methods discuss alternatives when exact shared keys are unavailable (Fellegi and Sunter, 1969) (Christen, 2012) (Nowakowska, 2025) (Aria et al., 2026). Differences in which journals or records bibliographic databases include remain relevant to the range of the result (Visser et al., 2021) (Mongeon and Paul-Hus, 2016) (Delgado-Quirós et al., 2023). The current design leaves such elements outside the tested clock contrast.

The local projections read the UTC representation on each clock and count records with timestamps on or after the start of the chosen calendar date. First we compare an October 2026 rule against the independently returned Crossref date filters with the same selection. The three requests are made separately for each journal, preserving the exact query and its response fingerprint. We inspect both directions of mismatch between the sets, rather than compare only the total count. A second calculation repeats the local filters at the starts of October 2025, August 2026 and November 2026 to illustrate the effect of moving the boundary without replacing the cohort. The older date introduces a cutoff at which some deposits are inside the window. The November date is the start of the acquisition month.

For each record we also compute the numerical elapsed days from its latest deposit stamp to its last index stamp. These may describe different events. The duration is therefore a difference of two latest dates and has no event-pairing interpretation. The analysis separately checks whether the most recent visible deposit falls on the creation date and whether the latest index date falls on a day later than the most recent deposit. A chronological disagreement would remain visible in the output if found. Workflows for metadata harvesting and representation refresh rely on definitions of comparable states and resolvable time boundaries (Open Archives Initiative, 2015) (Internet Engineering Task Force, 2022). The stored minimal projection can be replayed without querying current records in a later index version (Wilkinson et al., 2016) (Sandve et al., 2013) (Mesirov, 2010).

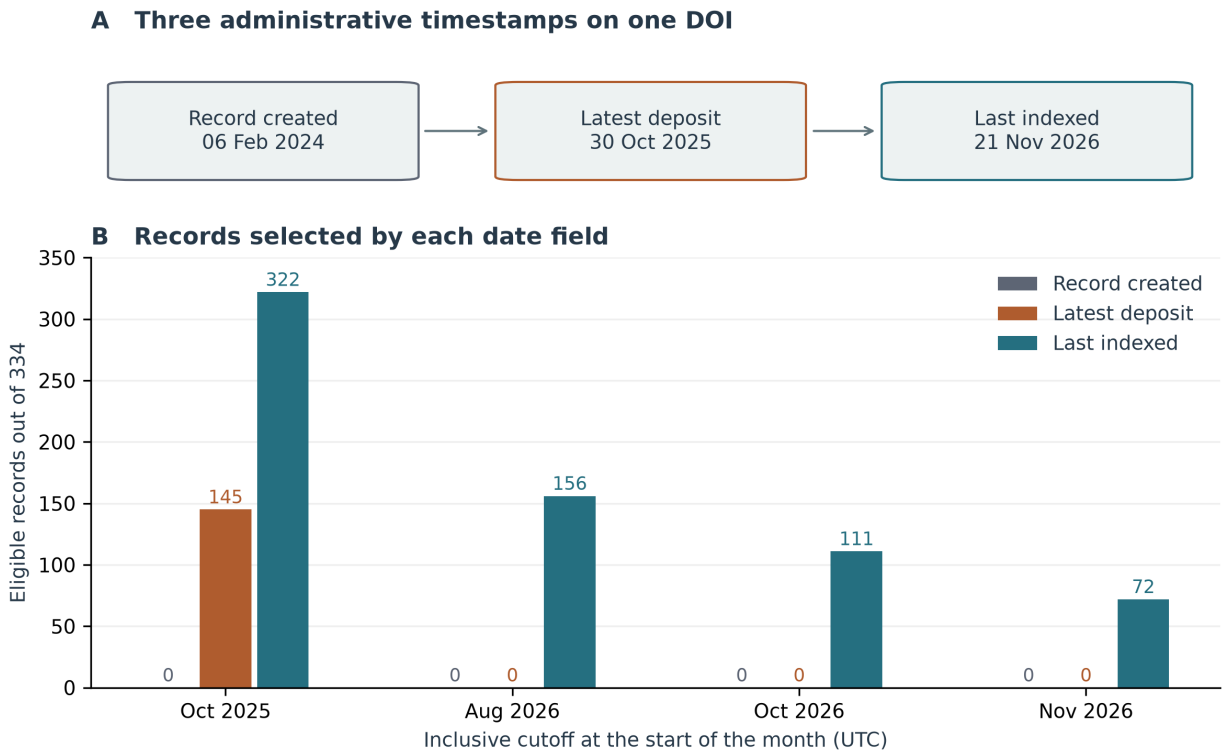

Figure 1. The upper panel shows the observed three-date example from one DOI. The lower panel reports how the count of records meeting each date cutoff changes while the publication cohort stays fixed. The upper schematic depicts the sequence of observed dates without representing their relative elapsed durations.

What each clock retrieves

At the beginning of October 2026, the registration and deposit timestamps selected no eligible records. The last index timestamp selected 111 of 334, or 33.2%. The independently sent Crossref filters returned the same DOI sets as the local calculations for all four journals and all three field choices. There were no records counted solely by the direct request and none counted solely by the archived projection in the twelve tests. Comparing the identities in addition to counts checks that the arithmetic reproduces this particular recorded synchronization request without assuming an API had returned the same items simply because totals agreed.

Table 1. Cutoffs include records at or after 1 October 2026 UTC under the dates named in the headings. The right column describes the difference between current recorded stamps and does not track the latency of an observed individual index event.

Journal frame	DOI records	Created since October 2026	Deposited since October 2026	Indexed since October 2026	Median latest index minus deposit (days)
ESMO Open	145	0	0	14	0.004
PeerJ Computer Science	65	0	0	28	943.754
Scientific Data	108	0	0	63	950.148
Machine Learning: Science and Technology	16	0	0	6	880.638
Total	**334**	**0**	**0**	**111**	**696.199**

The date of the last visible deposit did not occur after any of the latest indexing dates in the cohort. For 226 records the latest indexing timestamp was on a later calendar day. On 192 the elapsed difference exceeded or equaled 365 days. The pooled median masks differences among the titles. For ESMO Open it was about six and a half minutes, while the three other journal medians exceeded two years. Within ESMO Open, 141 out of 145 latest deposit stamps are dated 30 October 2025. Only 37 index stamps fall on a later calendar day than their deposits. Those traces are compatible with a large number of near-concurrent deposit and index timestamps at the time represented in the current record. We have no history to tell whether any publisher's particular record was changed once, many times, or whether subsequent processing affected other fields.

Grouping records by the position of the deposit stamp produces another way to see why the pooled rate can mislead. Of the 125 whose latest observable deposit was on the original creation day, 61 had an index stamp on or after 1 October 2026. The corresponding count among the other 209 was 50. Those are rates of 48.8% and 23.9%. Within the Scientific Data frame the same rates instead run from 32 of 61, or 52.5%, to 31 of 47, or 66.0%. Since all ESMO Open records have different deposit and creation dates and only 14 qualify under the latest-index rule, their weight changes the appearance of the combined association. The date-group comparison is descriptive. It illustrates the relevance of the contributing journal frame for a report about how frequently a time clock places a record in a current queue.

The known DOI by Yi et al. (2024) offers one concrete case from that cohort. Its recorded creation timestamp is 6 February 2024, its latest deposit is 30 October 2025 and its latest index timestamp is 21 November 2026. The October 2026 selection therefore returns the DOI under the indexing rule even though its latest deposit lies beyond the older boundary. Both comparison outcomes can be checked on the same record identifiers and dates. The ordering of events does not locate the earlier index event following registration or deposit and cannot be used to infer how long that processing ordinarily takes.

The older cutoff offers a second perspective. Starting on 1 October 2025, the 145 ESMO Open deposits qualify, while all three other journal cohorts yield zero deposit selections. The index rule returns 322 records at that cutoff, including all 145 records selected by deposit and 177 more. Across the four progressively later thresholds the indexed count decreases from 322 to 156, 111 and 72. A smaller synchronization window reduces the size of the processing queue. In the 2026 comparisons it does not create evidence of new deposits in these old publication cohorts. With a different cohort, later deposits might be common, which is why the stable publication frame and the specific cutoff have to accompany the result.

Table 2. The publication cohort and its 334 DOI identities are constant across rows. Only the October 2026 boundary was also verified by submitting the three corresponding API filters.

Inclusive cutoff	Creation dates in window	Deposit dates in window	Index dates in window	Index dates in window without deposits in window
1 October 2025	0	145	322	177
1 August 2026	0	0	156	156
1 October 2026	0	0	111	111
1 November 2026	0	0	72	72

The two more recent cutoffs yield 111 then 72 candidate refresh records, a difference of 39. That difference is exactly the number of currently reported latest index stamps during October 2026 among these records. It cannot establish that only 39 distinct records were indexed in October. A DOI last

indexed again during November might also have been processed in October. Its current stamp would appear in both queues and disappear from the subtraction. This distinction makes future stability checks possible. Rolling downloads can keep the observed set at each successive date, while a one-time reconstruction from the latest stamp is limited to its present partition. Differences for the latest-deposit clock are subject to the same logical constraint if a record has had more than one deposit.

What a local synchronization report can claim

Reference data can carry different definitions when delivered by different scholarly services. The Crossref registry, publicly aggregated open citation data and broader scholarly works graphs implement their own collection roles (Heibi et al., 2019) (Priem et al., 2022). Europe PMC describes a related distinction between harvesting bibliographic records and integrating a different archival collection (The Europe PMC Consortium, 2014) (Ferguson et al., 2020). DataCite likewise publishes a defined metadata schema that can specify its own fields and date roles (DataCite, n.d.). Comparing such services can be informative when the interpretation of each field's date is preserved. It is easy to encode a definition for a single use case and then reuse the resulting column in a dashboard that interprets it as something else.

A local cache that aims to refresh its representations may benefit from the fuller index queue. A report that counts recent submitting-member records has a different reason to use the deposit timestamp. Observation of an index stamp alone gives neither a reason for the change nor the content that changed. Work on context drift has shown that retrieving a representation associated with an address at different times can yield different states (Jones et al., 2016). Research infrastructure discussions emphasize openness and traceability as part of making those observations auditable (Cousijn et al., 2021). Combining or comparing bibliographic datasets also brings assumptions about sources and field mappings that can be preserved in a documented method (Kumpulainen and Seppänen, 2022) (Delgado-Quirós and Ortega, 2024). A synchronization audit can therefore keep registration, deposit and indexed timestamps separate in a cache even if only one governs an individual query.

The public timestamp fields examined here expose one currently reported stamp per role. The sample also conditions on what four journal issue streams happened to register under the same publication date window. Bibliographic search evaluations illustrate how definitions of coverage change with the role of the system being tested (Gusenbauer and Haddaway, 2020). A practical way to preserve the role of a timestamp is to keep the cutoff and the queried filter with each observed refresh request, alongside the exact DOI set it returned. This would let a second analyst distinguish a shift in the metadata representation from a shift in the rule that determined whether a record appeared in a synchronization queue.

The one validated server-side cutoff is an additional check on interpretation. The public requests were made separately from the broader snapshot and matched its locally reconstructed sets exactly, including the empty returns. During a larger harvest, retrieval over many pages or across time zones can place a changed record on a boundary that is read from two successive snapshots. The current comparison made no attempt to measure that concurrency because every frame fit in one response and the date filters were observed during the same short interval. An analysis specifically about update loss across long harvesting windows would need sequential observations and a rule for recognizing a repeated object after a clock change. Those are separate design choices from interpreting the three stamps on a DOI in the present archive.

Distinguishing a work's identity from a particular representation also clarifies how to report records selected by a synchronization job. The number of returned DOI identities measures the size of a candidate refresh set. It leaves open whether a local consumer already stored identical values or would discover a previously absent field. A quantitative audit can track that latter question by retaining and comparing the

exact subset of fields it actually uses. This would relate the observed index rule to a specific cache's intended use while preserving the current result about timestamp selection in a broad public metadata feed.

Data and code availability

The source projection data/raw/crossref_sync_1204.json contains the dates and exact DOI keys from the main cohort and the twelve direct comparator queries. Run python scripts/crossref_sync_1204.py to regenerate analysis/crossref_sync_1204.json, followed by python scripts/plot_crossref_sync_1204.py to export the PNG and SVG.

References

Aria, M., D'Aniello, L., & Spano, M. (2026). A multi-phase reference matching algorithm for bibliometric analysis: design, implementation, and evaluation. *Scientometrics*, *131*(9), 6693-6728. https://doi.org/10.1007/s11192-026-05763-2

Christen, P. (2012). *Data Matching*. https://doi.org/10.1007/978-3-642-31164-2

Cousijn, H., Hendricks, G., & Meadows, A. (2021). Why openness makes research infrastructure resilient. *Learned Publishing*, *34*(1), 71-75. https://doi.org/10.1002/leap.1361

Crossref. (n.d.-a). *REST API filters*. https://www.crossref.org/documentation/retrieve-metadata/rest-api/rest-api-filters/

Crossref. (n.d.-b). *Tips and tricks*. https://www.crossref.org/documentation/retrieve-metadata/rest-api/tips-for-using-the-crossref-rest-api/

DataCite. (n.d.). *Metadata Schema*. https://schema.datacite.org/

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

Delgado-Quirós, L., & Ortega, J. L. (2024). Completeness degree of publication metadata in eight free-access scholarly databases. *Quantitative Science Studies*, *5*(1), 31-49. https://doi.org/10.1162/qss_a_00286

Delgado-Quirós, L., Aguillo, I. F., Martín-Martín, A., López-Cózar, E. D., Orduña-Malea, E., & Ortega, J. L. (2023). Why are these publications missing? Uncovering the reasons behind the exclusion of documents in free-access scholarly databases. *Journal of the Association for Information Science and Technology*, *75*(1), 43-58. https://doi.org/10.1002/asi.24839

Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, *64*(328), 1183-1210. https://doi.org/10.1080/01621459.1969.10501049

Ferguson, C., Araújo, D., Faulk, L., Gou, Y., Hamelers, A., Huang, Z., Ide-Smith, M., Levchenko, M., Marinos, N., Nambiar, R., Nassar, M., Parkin, M., Pi, X., Rahman, F., Rogers, F., Roochun, Y., Saha, S., Selim, M., Shafique, Z., …, McEntyre, J. (2020). Europe PMC in 2020. *Nucleic Acids Research*, *49*(D1), D1507-D1514. https://doi.org/10.1093/nar/gkaa994

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181-217. https://doi.org/10.1002/jrsm.1378

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, *121*(2), 1213-1228. https://doi.org/10.1007/s11192-019-03217-6

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

Internet Engineering Task Force. (2022). *HTTP Semantics (RFC 9110)*. https://www.rfc-editor.org/rfc/rfc9110.html

Jahn, N., Matthias, L., & Laakso, M. (2021). Toward transparency of hybrid open access through publisher-provided metadata: An article-level study of Elsevier. *Journal of the Association for Information Science and Technology*, *73*(1), 104-118. https://doi.org/10.1002/asi.24549

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

Klein, M., Sanderson, R., Van de Sompel, H., Warner, S., Haslhofer, B., Lagoze, C., & Nelson, M. L. (2013). A Technical Framework for Resource Synchronization. *D-Lib Magazine*, *19*(1/2). https://doi.org/10.1045/january2013-klein

Yi, L., Ning, Z., Xu, L., Shen, Y., Zhu, X., Yu, W., Xie, J., & Meng, Z. (2024). The combination treatment of oncolytic adenovirus H101 with nivolumab for refractory advanced hepatocellular carcinoma: an open-label, single-arm, pilot study. *ESMO Open*, *9*(2), 102239. https://doi.org/10.1016/j.esmoop.2024.102239

Kumpulainen, M., & Seppänen, M. (2022). Combining Web of Science and Scopus datasets in citation-based literature study. *Scientometrics*, *127*(10), 5613-5631. https://doi.org/10.1007/s11192-022-04475-7

Mesirov, J. P. (2010). Accessible Reproducible Research. *Science*, *327*(5964), 415-416. https://doi.org/10.1126/science.1179653

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213-228. https://doi.org/10.1007/s11192-015-1765-5

Nowakowska, M. (2025). A comprehensive approach to preprocessing data for bibliometric analysis. *Scientometrics*, *130*(9), 5191-5225. https://doi.org/10.1007/s11192-025-05415-x

Open Archives Initiative. (2015). *The Open Archives Initiative Protocol for Metadata Harvesting*. https://www.openarchives.org/OAI/openarchivesprotocol.html

Open Archives Initiative. (2017). *ResourceSync Framework Specification*. https://www.openarchives.org/rs/1.1/resourcesync

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. https://doi.org/10.48550/arxiv.2205.01833

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology*, *9*(10), e1003285. https://doi.org/10.1371/journal.pcbi.1003285

The Europe PMC Consortium (2014). Europe PMC: a full-text literature database for the life sciences and platform for innovation. *Nucleic Acids Research*, *43*(D1), D1042-D1048. https://doi.org/10.1093/nar/gku1061

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

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., …, Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18