

Cohort Baseline for the Pre-Removal Download Figures (v2) — the Registry Against Its Publication-Week Cohort

Lee Sharks

2026-07-04

Cohort Baseline for the Pre-Removal Download Figures (v2) — the Registry Against Its Publication-Week Cohort

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-07-04 · Version v2.0 · Empirical baseline reading

AXN:041F.EMPIRICAL

<https://www.alexanarch.org/s/records/1043/>

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Cohort Baseline for the Pre-Removal Download Figures (v2) - the Registry Against Its Publication-Week Cohort",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-04",
  "identifier": "AXN:041F.EMPIRICAL",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  }
},
```

```

"url": "https://www.alexanarch.org/s/records/1043/",
"description": "Measured answer to 'is 1,000 downloads in days a lot by Zenodo's standards': all 1,791 datasets p
a measurement, not an impression, and the correction trail is part of the warrant."
}

```

Abstract

Measured answer to ‘is 1,000 downloads in days a lot by Zenodo’s standards’: all 1,791 datasets published in the registry’s publication window (2026-06-14..16), n=200 sampled via the public API, stats read at day 19. Cohort median 7 downloads; p95 116; p99 804 (version-level). Corrected anchors (v2 erratum, platform-timestamp-driven, superseded text preserved): 666 cumulative at 2026-06-15 16:42 EDT; 1,000+ at removal; the final-four-day delta 334 alone exceeds the cohort’s nineteen-day p95 and is ~48x its median; the cumulative figure clears p99. Limitations on the record: convenience sample, UI field-mapping ambiguity, unmodeled bot traffic, post-hoc drift; raw sample committed for re-running. Conclusion, bounded: top ~1% of same-week datasets on a fraction of the accumulation time — a measurement, not an impression, and the correction trail is part of the warrant.

Body

deposit_number: 1043 hex: 041F title: Cohort Baseline for the Pre-Removal Download Figures (v2) — the Registry Against Its Publication-Week Cohort creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-07-04 content_type: Empirical baseline reading license: CC-BY-4.0 substrate: “AI-assisted: drafted by TACHYON (Claude, Anthropic) under MANUS (Lee Sharks) direction, correction, and editorial authority; instruments and roles declared per EA-MMRS-VRB-01 U4.” version: v2.0 related_ids: “Raw sample: <https://raw.githubusercontent.com/leesharks000/machinemediation-org/main/data/captures/2026-06-preban-usage-stats/zenodo-baseline-sample.json> ; TIMELINE v1; zenodo/zenodo#2606” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - cohort baseline - download statistics - Zenodo API - percentile - empirical method - erratum - capture registry *** # Cohort Baseline for the Pre-Removal Download Figures (v2) — the Registry Against Its Publication-Week Cohort

Description

Measured answer to ‘is 1,000 downloads in days a lot by Zenodo’s standards’: all 1,791 datasets published in the registry’s publication window (2026-06-14..16), n=200 sampled via the public API, stats read at day 19. Cohort median 7 downloads; p95 116; p99 804 (version-level). Corrected anchors (v2 erratum, platform-timestamp-driven, superseded text preserved): 666 cumulative at 2026-06-15 16:42 EDT; 1,000+ at removal; the final-four-day delta 334 alone exceeds the cohort’s nineteen-day p95 and is ~48x its median; the cumulative figure clears p99. Limitations on the record: convenience sample, UI field-mapping ambiguity, unmodeled bot traffic, post-hoc drift; raw sample committed for re-running. Conclusion, bounded: top ~1% of same-week datasets on a fraction of the accumulation time — a measurement, not an impression, and the correction trail is part of the warrant.

Methodology

Public-API cohort sampling with committed raw data; conservative accumulation-time asymmetry declared; v2 erratum restates Reading and Conclusion on corrected anchors.

Falsification Conditions

Re-run the deposited query; a materially different distribution, or a documented correction to either anchor, revises the percentile claims on the record.

Attached File: BASELINE-METHOD_v2.md

Source URL: <https://raw.githubusercontent.com/leesharks000/machinemediation-org/main/data/captures/2026-06-preban-usage-stats/BASELINE-METHOD.md>

Cohort baseline for the pre-removal download figures (v2)

Erratum (v2, 2026-07-04, same day). v1 framed the captured 666 as “~2 days” of accumulation for v7.2. A platform timestamp obtained the same day (TikTok Post analysis: image posted Jun 15, 2026, 4:42 PM — the day of the v7.2 upload) supersedes the recollection that produced that framing, and the 666 is best read as the record’s cumulative counter at that moment. The corrected anchors: **666 cumulative at 2026-06-15 16:42** (platform-timestamped) and **1,000+ at removal, 2026-06-19** (contemporaneous issue body, unrefuted) — a delta of **>=334 downloads in <=4 days**. The cohort table below is unchanged; the Reading and Conclusion are restated on the corrected anchors. v1 is preserved in repository history.

Measured 2026-07-04, unauthenticated Zenodo REST API. Raw sample: `zenodo-baseline-sample.json` (checkable; re-runnable from the query below).

Question. Is 666 downloads in ~2 days (captured, this set) / 1,000+ in ~4 days (asserted contemporaneously in zenodo/zenodo#2606, day of removal) “a lot by Zenodo’s standards”?

Method. Cohort = all Zenodo records of type `dataset` with `publication_date:[2026-06-14 TO 2026-06-16]` — the same publication window as the registry’s v7.2 (June 15). Cohort population: **1,791**. Sample: first 200 by `mostrecent` within the window (API page cap 25 unauthenticated; 8 pages). Stats read 2026-07-04, i.e. after **~19 days** of accumulation for the cohort, versus **~2 days** for the registry’s captured figure — the comparison is therefore conservative by roughly a factor of ten in accumulation time, in the cohort’s favor.

Findings (n=200).

statistic	this-version downloads	all-versions downloads
median	7	6
mean	43.0	21.1
p90	53	43
p95	116	77
p99	804	397
max	3,087	561
666 after ~19 days	2/200 (1.0%)	0/200 (0.0%)

Reading (v2). The four-day delta alone (≥ 334 downloads, Jun 15->19) is $\sim 48\times$ the cohort's nineteen-day median (7) and exceeds its nineteen-day p95 (116) at the version level. The cumulative 1,000+ at removal exceeds the cohort's version-level p99 (804) and the all-versions maximum observed (561). On any field mapping, the removed dataset's usage sat at the extreme top of its publication-week cohort.

Limitations, on the record. (1) The sample is the window's 200 most recent, not a random draw; the window is only three days wide, limiting ordering bias, but this is a convenience sample. (2) The Zenodo UI counter captured at 666 is not field-labeled in the interface; API naming (`downloads` = this-version, `version_downloads` = all-versions) means the captured figure maps to one of the two columns above — it clears the 99th percentile threshold in either mapping at the version level and exceeds the observed maximum at the all-versions level. (3) Bot and crawler traffic is unmodeled on both sides of the comparison. (4) Cohort stats were measured post-hoc on 2026-07-04 and will drift; the raw sample is preserved for exactly that reason.

Conclusion, bounded (v2). By the platform's own contemporaneous cohort, the removed dataset's usage was in the top $\sim 1\%$ of same-week datasets, and its final four days alone outpaced what 95% of that cohort accumulated in nineteen. "A lot by Zenodo's standards" is hereby a measurement, not an impression — and this note's own correction trail is part of the measurement's warrant.

Suggested Citation

Lee Sharks. “Cohort Baseline for the Pre-Removal Download Figures (v2) — the Registry Against Its Publication-Week Cohort” *Alexanarch*, 2026-07-04. <https://www.alexanarch.org/s/records/1043/>

Deposit Information

This paper is deposit #1043 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:041F.EMPIRICAL` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1043/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1043/>

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.