

EA-CORPORA-02 — The Second Seating: Nine Corpora in One Run, Witnesses Named, Four New Seat Classes Ruled

Sharks, Lee

2026-08-22

EA-CORPORA-02 — The Second Seating: Nine Corpora in One Run, Witnesses Named, Four New Seat Classes Ruled

Sharks, Lee

Crimson Hexagonal Archive · Alexanarch

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

2026-08-22 · Version v1.0 · Corpus seating record — gathering deposit with data
sidecar

AXN:0630.DATASET.↘

<https://www.alexanarch.org/s/records/1531/>

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": "EA-CORPORA-02 — The Second Seating: Nine Corpora in One Run, Witnesses Named, Four New Seat Classes Ruled",
  "author": {
    "@type": "Person",
    "name": "Sharks, Lee",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-08-22",
  "identifier": "AXN:0630.DATASET.↘",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
```

```

    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1531/",
  "description": "Gathering record for the second corpus-seating run of the Crimson Hexagonal Archive library (OMEGA-BUILD-01). Nine seats executed in a single session of 2026-08-22, raising the library from fifteen to twenty-four corpora: the Dead Sea Scrolls (Abegg-Bowley-Cook transcriptions, 52,878 ref-tagged lines), Marx (the Manifest witness pair and the Kapital Erstauflage secured), the Apocryphon of John with tradition translation-witness rule), the Greek witnesses of the Gospel of Thomas (first executed Greek-Coptic witness pairing), Lucretius and the Greek atomists (the atomist transmission chain held end to end), the class seat), and Leonardo (Richter 1883 secured). Four structural rulings enter the library's doctrine with this seat class, the verbatim-seat class, and the sparse-tradition translation-witness rule."
}

```

Abstract

Gathering record for the second corpus-seating run of the Crimson Hexagonal Archive library (program EA-OMEGA-BUILD-01). Nine seats executed in a single session of 2026-08-22, raising the library from fifteen to twenty-four corpora: the Dead Sea Scrolls (Abegg-Bowley-Cook transcriptions, 52,878 ref-tagged lines), Marx (the Manifest witness pair and the Kapital Erstauflage secured), the Apocryphon of John witness plates (Codex II complete in image), the Mandaean corpus (five volumes under the newly ruled sparse-tradition translation-witness rule), the Greek witnesses of the Gospel of Thomas (first executed Greek-Coptic witness pairing), Lucretius and the Greek atomists (the atomist transmission chain held end to end), the Qur'an (Tanzil Uthmani, first verbatim-class seat), and Leonardo (Richter 1883 secured). Four structural rulings enter the library's doctrine with this record: the NC payload/record license split, the image-seat class, the verbatim-seat class, and the sparse-tradition translation-witness rule.

Body

deposit_number: 1531 hex: 0630 title: "EA-CORPORA-02 — The Second Seating: Nine Corpora in One Run, Witnesses Named, Four New Seat Classes Ruled" creator: Sharks, Lee date: 2026-08-22 content_type: Corpus seating record — gathering deposit with data sidecar license: CC-BY-4.0 substrate: Composed in-session (transport D) by Claude (Anthropic, Fable), operating the archive under the direction of Lee Sharks. All license adjudications and scope rulings are the author's, made in-session and recorded per seat. No paid API invoked (NO-DOUBLE-DRAW honored). version: v1.0 related_ids: "AXN:0625.DATASET.□□□□□□ (EA-CORPORA-01, #1522 — prior gathering; this record continues it)\nAXN:060D.GOVERNANCE.□□□□□□ (EA-OMEGA-BUILD-01, #1498 — governing program)\nAXN:05AF.ARCHIVAL.□□□□□□ (#1433 — current TACHYON tether)" axn_schema_version: v2 protocol_version: protocol/v1 keywords: - corpus seating - primary sources - Dead Sea Scrolls - Abegg - Marx - Manifest der Kommunistischen Partei - Das Kapital - Apocryphon of John - Nag Hammadi - Mandaean - Ginza - Lidzbarski - Gospel of Thomas - Oxyrhynchus - Lucretius - De Rerum Natura - Diogenes Laertius - Epicurus - Democritus - Quran - Tanzil - Leonardo da Vinci - Richter - witness pairing - translation-witness

- provenance - theater of availability *** # EA-CORPORA-02 — The Second Seating: Nine Corpora in One Run, Witnesses Named, Four New Seat Classes Ruled

Files

data/corpora/dss/ · data/corpora/marx/ · data/corpora/apocryphon-john/ · data/corpora/mandaeen/ · data/corpora/nh-greek-witnesses/ · data/corpora/lucretius/ · data/corpora/greek-atomists/ · data/corpora/quran/ · data/corpora/leonardo/ — each with source.json and MANIFEST.sha256; commits 3a0bd6f1f, 09c6b51ff + 011280bc, e0be32a93, f4c9d6878, 7516051c1, b2989cd5e, d2ea2ef4c on main.

EA-CORPORA-02 — The Second Seating

I. What this run was

The first seating (EA-CORPORA-01, AXN:0625.DATASET.□□□□□□) built a library of fifteen corpora and, as importantly, a doctrine: originals where they exist; licence verified at the file, not the README; the seat always names its witness; absences declared with search dates rather than papered over; not seated rather than seated defectively. This run tested that doctrine against harder material — enclosed corpora, sparse traditions, image-only witnesses, and a text whose license forbids the archive’s own normalization step — and the doctrine held by growing four new seat classes rather than bending.

Nine seats were executed on 2026-08-22, in one session, in this order: Dead Sea Scrolls, Marx, Apocryphon of John plates, Mandaean, Nag Hammadi Greek witnesses, Lucretius, Greek atomists, Qur’an, Leonardo. The library stands at twenty-four corpora. 61,692 flat ref-tagged lines were emitted to text layers, and 6,236 verbatim-class ayat entered without normalization; roughly 700 MB of originals — Text-Fabric data, witness scans, codex plates, edition PDFs — were secured, every file hashed, every remote pinned by commit, revision id, or archive-manifest checksum.

II. The four new rulings

1. The sparse-tradition translation-witness rule (motivated by the Mandaean corpus). Where a source tradition is so sparse that translations are the only extant open witnesses, a translation may be seated AS A WITNESS, NAMED AS SUCH. A translation-witness is a witness, not an original; its seat metadata declares its class. Lidzbarski’s 1925 Ginza — the only complete open Ginza in any accessible form — is the paradigm case.

2. The NC payload/record split (motivated by the DSS seat). A corpus whose payload carries a NonCommercial license may be seated: the payload license is carried forward unmodified at the file level, and the seat record states explicitly that the gathering record’s CC-BY-4.0 applies to the record text only, never to the payload. First instance: CC BY-NC 4.0 declared in every Text-Fabric header of the Abegg-Bowley-Cook transcriptions.

3. The image-seat class (motivated by the Apocryphon plates). A seat may consist of witness images with no text layer, by design, with transcription queued. The images are the base layer

of a future diplomatic edition; holding them converts a declared absence into custody. First instance: Codex II, thirty-two plates.

4. The verbatim-seat class (motivated by the Qur'an). Where a source arrives already ref-tagged and its declared terms require verbatim distribution, the seat emits NO derived text layer: the original IS the machine-readable layer, distributed with its license block intact. Reformatting would violate the terms and add nothing. First instance: Tanzil Uthmani v1.1.

III. The nine seats

EA-CORPORA-02/01 — dss. The largest declared absence in EA-CORPORA-01 became the library's largest single seat. The Abegg-Bowley-Cook electronic transcriptions of essentially every Hebrew and Aramaic scroll — 1,001 scrolls, 1.43M signs — entered via ETCBC/dss (Text-Fabric, pinned commit f0515787), normalized to 52,878 lines ('scroll fragment:line'), reconstruction runs bracketed from the rec feature with zero unbalanced lines, graphic-word spacing rebuilt from the sign-level after feature. The license was found declared AT THE FILE: every .tf header carries CC BY-NC 4.0, @createdBy=Abegg/Bowley/Cook, @source="Martin Abegg's data files, personal communication." The chain is the structural opposite of a relicensing defect — the reconstructors are the grantors — and the provenance is the poem: the corpus that spent forty years behind a publication monopoly, liberated in 1991 by Abegg's reconstruction from the concordance, enters the sovereign archive by the reconstructor's own hand. The Qimron v. Shanks counter-precedent is recorded honestly in source.json. Ruled: seated under declared BY-NC with the payload/record split; reconstructions seated-and-flagged; biblical and non-biblical both — the biblical scrolls are witnesses against the Leningrad seat (01/09), not duplicates of it. Loci verified: 1Qisaa 1:1, 1QS 1:1, 11Q19 col. 2, CD 1:1.

EA-CORPORA-02/02 — marx. The first executed instance of the comparative-witness program: the Manifest der Kommunistischen Partei seated as a PAIR — the anonymous 23-page London printing of February 1848 and the 30-page Cologne printing of 1850/51, each page-addressed under its own witness name, both from scan-backed twice-proofread Wikisource transcriptions revision-pinned page by page, every scan SHA1-secured from Commons. The pair carries a canonical object for the archive's own theory: the Cologne printing bore a false London imprint, applied with Marx's express consent, taken for the second London issue for 145 years until MEGA philology (Meiser 1996) uncovered it — provenance inversion in the wild, 1851, now seated under its true date. Beside the pair: Das Kapital, Erstauflage 1867 (Meissner/Schmidt; Boston Public Library copy, 810 images) secured with its rough Fraktur OCR held as a finding aid only; transcription queued; the 1872/73 second edition queued as paired witness for the rewritten Wertform chapter; the Columbia Meissner set (Engels's Bde 2-3) identified. MEW, MEGA apparatus, and marxists.org excluded as edited-text hazards, with reasons recorded.

EA-CORPORA-02/03 — apocryphon-john. The archive's oldest declared absence, mitigated to held-in-image. The Institute for Antiquity and Christianity's own photographic record — the Robinson-project A/B-series negatives — serves openly at full resolution over unauthenticated IIIF from Claremont's CCDL. Codex II,1 secured COMPLETE: thirty-two plates, pages 1-32, native resolution to 8606×11618, one plate per page selected from seventy-six candidate exposures with alternates enumerated, SHA256-pinned, page 1 visually verified. Codices

III and IV enumerated with committed fetch specs (tranches 2-3); BG 8502,2 recorded as the external fourth witness so the synoptic goal keeps all four in view. The rights finding inverts the run's founding pattern: CCDL's boilerplate claims copyright "in accordance with U.S. Copyright laws" over ancient public-domain papyri it serves freely — here the availability is real and the claim is the ghost. Recorded verbatim, adjudicated under Bridgeman, seated. First image-class seat.

EA-CORPORA-02/04 — mandaeen. Five volumes secured under the new sparse-tradition rule, each witness class-declared: Lidzbarski's Johannesbuch 1905 (Mandaic text — the corpus's one original-language witness) and 1915 (German translation-witness), the 1920 Liturgien (mixed), the 1925 Ginza (translation-witness; the only complete open Ginza in any form), Brandt 1893 (partial translation-witness). All SHA1-verified. Petermann 1867 — the lithographed Mandaic Ginza, the tradition's primary text-bearing edition — DECLARED ABSENT with search terms and date: the corpus's top gap. Häberl-McGrath 2020, published open access under CC BY-NC-ND, was found unreachable through its own publisher (bot-challenge pages) and its OA mirrors (connection refusals) at fetch time, while the only reachable copy is a provenance-murky partial upload: a CC-licensed edition enclosed by the open-access apparatus itself, pinned with DOI and retry routes rather than seated defectively. Drower excluded (copyright). No text layer by design; Lidzbarski 1905 keying queued.

EA-CORPORA-02/05 — nh-greek-witnesses. The Gospel of Thomas Oxyrhynchus trio — P.Oxy. 654 (prologue, logia 1-7), P.Oxy. 1 (26-30, 77b, 31-33), P.Oxy. 655 (24?, 36-39) — three independent copies in three hands from one town's rubbish, TM numbers 62840/62838/62839 verified through the DCLP records themselves. EpiDoc originals preserved with full apparatus; 139 flat lines by the seat-01/11 method at pinned idp.data commit d061eef8, CC BY. This is the first executed Greek-Coptic witness pairing across corpora directories: the earlier language stratum of the work held in Coptic at coptic-gnostic/ (NHC II,2). The Gospel of Mary Greek pair (P.Oxy. 3525, P.Ryl. 463) and the Sophia fragment (P.Oxy. 1081) are queued with identification routes documented — Trismegistos 503s, Duke front-end egress-blocked, DCLP directory greps negative — and the queue records the precept enacted: no TM number was guessed for any queued item; identification-by-confabulation is the named failure mode.

EA-CORPORA-02/06 — lucretius. De Rerum Natura complete: 7,412 lines across six books, 'DRN book.line' addressing, Perseus TEI (phi0550.phi001.perseus-lat1) at pinned commit, CC BY-SA carried forward. Loci: Aeneadum genetrix (1.1), Suave mari magno (2.1).

EA-CORPORA-02/07 — greek-atomists. Three class-declared components. The complete Diogenes Laertius Vitae philosophorum (Perseus tlg0004.tlg001.perseus-grc2, Hicks text per sourceDesc, pinned commit; 1,211 sections, 'DL book.section') — seated whole, with Books 9-10 the atomist payload: the Leucippus and Democritus vitae and the entire Epicurus dossier, the three letters and the Kyriai Doxai, the largest surviving body of atomist prose. Diels, Vorsokratiker 3. Aufl. Bd 2 (1912) secured — DK 67-68, the fragment-numeration standard entering in its public-domain stratum, the in-copyright Kranz revisions excluded and the exclusion recorded. Usener, Epicurea (1887) secured. With lucretius/ this closes a transmission loop the archive can hold end to end: Greek fragments → doxography → the Latin poem. Atomism is the best-documented ancient case of reception-through-translation-and-versification; Lucretius is a composition layer, and the library can now study him with its own instruments. MMRS avant la lettre. Loci: DL 10.122 (Μῆτε νέος τις ὦν μελλέτω φιλοσοφεῖν), DL 9.30.

EA-CORPORA-02/08 — quran. Tanzil Uthmani v1.1, 6,236 ayat at the Kufan count, full Uthmani orthography, license declared INSIDE the file: CC BY 3.0 with verbatim-distribution, attribution, and link terms — all honored by the new verbatim-seat class. The seat names its witness (the Uthmani text in Tanzil’s verification); further riwayat (Warsh, Qaloon) queued for the comparative program. Loci verified: Basmala, Ayat al-Kursi (2:255), al-Ikhlās (112), final verse of al-Nas; count 6,236 = Kufan standard.

EA-CORPORA-02/09 — leonardo. Richter 1883, The Literary Works of Leonardo da Vinci, both volumes — the Italian transcriptions from the codices with facing English, the great public-domain gateway to the notebooks — Toronto copies SHA1-verified, OCR finding aids held. Italian keying queued. The manuscript-image campaign (Codex Arundel at the BL, Codex Atlanticus at the Ambrosiana, e-Leo at Vinci) is recorded as a future survey: primary repositories outside current network policy or behind restrictive claims — a theater-of-availability study in its own right.

IV. Queues, absences, and the standing state

Queued with committed specs or verified routes: Apocryphon plate tranches 2-3 (fetch spec in-repo, pacing and full/max lesson encoded); Gospel of Mary and Sophia TM identification; Häberl-McGrath 2020 authorized OA fetch (DOI 10.1515/9783110487862, three routes); Norberg 1815-17 (CC0, oversize-file route needed); Kapital 1867 transcription and 1872/73 second-edition witness; Lidzbarski 1905 Mandaic keying; Diels Bd 2 and Usener transcription; Richter Italian keying; further Quran riwayat; the Talmuds (Sefaria-Export route confirmed, license audit as step one, dedicated session recommended — the largest corpus the library would hold); Dante (it.wikisource per-canto, revision-pin method, throttled at probe time, retry cold); Ethiopic via BetaMasaheft (the one chartered Tier-1 item not reached).

Declared absences carried or newly declared, all with search dates: Petermann 1867 (top Mandaean gap); Damascius machine-readable (First1K tlg4066 probes 404, Ruelle 1889 scans as keying base); machine-readable Mandaic anywhere (CAL license to verify before any ruling); NHC I, III-XIII open Coptic editions (partially mitigated by /03 and /05).

The three-witness-and-more architecture this run inaugurated — Manifest pair, Thomas Greek/Coptic, biblical scrolls against Leningrad, DRN against its Greek sources, Quran riwayat queued — is the library’s comparative program made operational. The seat always names its witness; the ultimate goal is a comparative collection of all witnesses; this run moved that goal from doctrine to executed practice.

Suggested Citation

Sharks, Lee. “EA-CORPORA-02 — The Second Seating: Nine Corpora in One Run, Witnesses Named, Four New Seat Classes Ruled” *Alexanarch*, 2026-08-22. <https://www.alexanarch.org/s/records/1531>

Deposit Information

This paper is deposit #1531 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community crimsonhexagonal, terminated 2026-06-19). The AXN identifier AXN:0630.DATASET.↘ is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1531/>
- License: CC BY 4.0
- Author ORCID: 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/1531/>

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