

The digitized Leibniz Nachlass by the numbers

A corpus census of the GWLB digital collections (Leibniz Legible, Phase A1)

Evan Tabak Atlas · ORCID 0009-0007-7374-2338

2026-09-16

doi:10.5281/zenodo.22782815 (this version)

License: CC BY 4.0

Generated from <https://github.com/marchofhaires/leibnizlegible> at commit a7c729c.

ABOUT THIS DOCUMENT

This is a working report of Leibniz Legible, an open, one-person project building an access layer for the digitized Leibniz Nachlass held by the Gottfried Wilhelm Leibniz Bibliothek (GWLB), Hannover: every page machine-transcribed with per-line confidence and provenance, searchable, and browsable in an open IIIF viewer linked to the scholarly catalogue. The register is deliberately under-claiming. What the project produces is machine output with honest labels, subordinate to the Akademie-Ausgabe, never an edition. The report below is reproduced unchanged from the project repository, where it is regenerated from the project's own data store by the command named in its header; the numbers are as measured on the dates stated in the text. Images are © none: the GWLB manuscript scans carry the Public Domain Mark 1.0 and are never rehosted by the project; catalogue data is from the Leibniz-Katalog (BBAW/TELOTA, CC BY 4.0); the HTR model and ground truth used are from the ERC project PHILIUMM (CC BY 4.0), cited inside.

What this report is: the first published page-image count of the Hannover Leibniz Nachlass, computed from the library's own OAI-PMH/METS metadata (harvest of July 2026) and cross-checked against IIIF manifests: 2,225 works and 236,795 page images across the five Leibniz sets, with the per-set breakdown, the pages-per-work distribution, shelfmark coverage, the IIIF-versus-static-JPEG delivery split, anomalies, and, appended from Phase A2, the statistics of the full image pull (236,779 pages, 395.6 GB) that the transcription pipeline ran on.

Corpus census — digitized Leibniz Nachlass (GWLH Hannover)

Leibniz Legible, Phase A1. Generated 2026-07-29T00:44:33Z from the OAI-PMH harvest of the five Leibniz sets at `digitale-sammlungen.gwlh.de`.

This is, to our knowledge, the **first published page-image count** of the Hannover Leibniz Nachlass computed from the library's own metadata. Each work's page count is the number of `TYPE="page"` divisions in its METS *physical* structMap — equal to its IIIF manifest's canvas count (verified on object 00068642: 4 = 4), so the total needs no image download. Numbers will drift as the GWLB re-exports; regenerate with `leibniz harvest census`.

Headline

- **Unique works (objects):** 2,225
- **Unique page images:** 236,795
- **Gate:** the count is within the SPECS §1 expected band (150,000–250,000 images).

Works and pages per set

Each work is assigned to one **primary set** (priority: Handschriften → Briefwechsel → Marginalien → Rekonstruktionen → Leibnitiana), so these rows are disjoint and sum to the unique totals. The OAI sets themselves overlap (next table).

Primary set	Works	Page images	% of pages
LeibnizHandschriften	756	58,828	24.8%
LeibnizBriefwechsel	1,059	72,284	30.5%
LeibnizMarginalien	396	103,887	43.9%
leibniz-rekonstruktionen	2	28	0.0%
Leibnitiana	12	1,768	0.7%
Total (unique)	2,225	236,795	100%

OAI set membership (overlap cross-check)

A work counts once per set it belongs to here, so columns overlap and sum to more than the unique total. The **OAI size** column is the endpoint's own `completeListSize` (verified live 2026-07-28); **membership** is what we parsed and retained — they should match set-by-set.

OAI set	OAI size	Works (membership)	Works (primary)
LeibnizHandschriften	756	756	756
LeibnizBriefwechsel	1,060	1,060	1,059
LeibnizMarginalien	396	396	396
leibniz-rekonstruktionen	2	2	2
Leibnitiana	1,823	1,823	12

1,811 of 2,225 works (81.4%) belong to more than one Leibniz set — the reason a naive sum of set sizes (4,037) overcounts the 2,225 real objects.

Image delivery: IIIF vs. static JPEG

A finding that revises SPECS §1.1 (which assumed IIIF Image API tiles from pyramid TIFFs for the whole Nachlass): **only part of the corpus is served via IIIF**. Works whose METS carries a `mods:identifier[@type='iiif']` have a Presentation manifest and an Image API service (`.../iiif/{id}/ptif/...`); the rest 302-redirect `.../manifest.json` to a viewer page and expose only static JPEG derivatives at `.../content/{id}/jpgs/...`. Spot-checks (~9 works) confirm the flag tracks real availability closely, though not perfectly (one unflagged object also served a manifest), so treat this as the lower bound on IIIF coverage; A2 establishes per-work delivery definitively.

- **IIIF-served works:** 812 (36.5%), 77,633 pages (32.8%).
- **Static-JPEG-only works:** 1,413 (63.5%), 159,162 pages (67.2%).

Primary set	IIIF works	Static-only works	IIIF pages	Static pages
LeibnizHandschriften	750	6	55,593	3,235
LeibnizBriefwechsel	0	1,059	0	72,284
LeibnizMarginalien	60	336	21,832	82,055
leibniz-rekonstruktionen	0	2	0	28
Leibnitiana	2	10	208	1,560

Implication: A2's image cache cannot use the IIIF Image API for the majority of pages — it must fall back to the static JPEG derivative; D2's deep-zoom viewer likewise degrades to plain images where no Image API exists.

Pages per work

- Min 0 · Max 3,572 · Mean 106.4 · Median 26
- Quartiles (Q1/Q2/Q3): 8 / 26 / 94
- Works with ≥ 1 canvas: 2,198 of 2,225

Pages per work	Works
0 (no canvases)	27
1	0
2	88
3–5	234
6–10	347
11–20	316
21–50	409
51–100	264
101–200	196
201–500	219
501+	125

Largest works

Object id	Primary set	Pages	Shelfmark
DE-611-HS-3618673	LeibnizBriefwechsel	3,572	LBr. F 20
DE-611-HS-3616805	LeibnizBriefwechsel	1,752	LBr. 79
00068377	LeibnizHandschriften	1,668	LH 1, 19
1686254318	LeibnizMarginalien	1,310	Leibn. Marg. 64:1
1815201908	LeibnizMarginalien	1,294	Leibn. Marg. 230, Stück 1
DE-611-HS-960122	LeibnizBriefwechsel	1,262	LBr. 228
867984570	LeibnizMarginalien	1,209	Leibn. Marg. 100:3
DE-611-HS-974102	LeibnizBriefwechsel	1,160	LBr. 411
00068812	LeibnizHandschriften	1,106	LH 23, 2, 21
851387705	LeibnizMarginalien	1,089	Leibn. Marg. 100:1/2, Stück 1
1759045683	LeibnizMarginalien	1,064	Leibn. Marg. 121
876358024	LeibnizMarginalien	1,063	Leibn. Marg. 116
DE-611-HS-860518	LeibnizBriefwechsel	1,058	LBr. 676
DE-611-HS-4280474	LeibnizHandschriften	1,045	Ms IV, 471 : M-S
79876922X	LeibnizMarginalien	1,029	Leibn. Marg. 65

Shelfmark coverage

How many works carry a parseable Leibniz signature (LH / LBr / Leibn. Marg. / LK). **other** = has a shelfmark that doesn't match those families; **none** = no **shelfLocator** at all. (Robust normalisation is Phase A3; this is the coarse family classifier.)

Set	Marg	LBr	LK	LH	other	none	parseable %
LeibnizHandschriften	0	1	0	750	4	1	99.3%
LeibnizBriefwechsel	0	1,058	0	0	1	0	99.9%
LeibnizMarginalien	370	0	0	0	4	22	93.4%
leibniz-rekonstruktionen	0	0	0	0	1	1	0.0%
Leibnitiana	0	0	0	5	7	0	41.7%
All	370	1,059	0	755	17	24	98.2%

Anomalies

- **Zero-canvas works:** 27 (1.2%). These are container/anchor records with no physical pages (e.g. multi-part parents).
- **1040716245** (LeibnizMarginalien) — La Vie De Monsieur Des-Cartes
- **1049318668** (LeibnizMarginalien) — Machines Nouvellement Executees, Et En Partie Inventees Par le sieur Hubin, Ema...
- **1663593558** (LeibnizMarginalien) — Trophees tant sacres que profanes De La Dvché De Brabant
- **1671913108** (LeibnizMarginalien) — [Theologia Moralis, Adversus Laxiores Probabilistas] Theologia Moralis, Adversv...

- 1677984635 (LeibnizMarginalien) — Exercitiorvm|| Logicorum|| Andreae Li-||bavii M. D. P.|| Liber ...||
- 1691256005 (LeibnizMarginalien) — Alamannicarum Rerum Scriptores Aliquot Vetusti, a quibus Alamannorum, Qui Nunc ...
- 1759056197 (LeibnizMarginalien) — Geometria
- 769493815 (LeibnizMarginalien) — Entretiens Sur La Metaphysique, Sur La Religion, Et Sur La Mort
- **Works with no shelfmark:** 24 (1.1%).
- **Shelfmark strings shared by >1 work:** 56 (exact-string; variant normalisations counted separately).
- ZEN Leibn. Marg. 41 → 29 works (798250313, 798251158, 798278285, 798280190...)
- ZEN Leibn. Marg. 42 → 20 works (799210226, 799214043, 799215783, 799264644...)
- ZEN Leibn. Marg. 0,1 → 15 works (797486364, 797487301, 797488162, 79749524X...)
- ZEN Leibn. Marg. 131 → 11 works (816171491, 816176299, 816218927, 816552622...)
- Leibn. Marg. 72 → 7 works (818661070, 818663391, 818669012, 818670150...)
- Leibn. Marg. 113 → 7 works (862386217, 862388155, 862388503, 862388961...)
- Leibn. Marg. 128 → 5 works (1032725095, 1032726466, 1032770724, 1032771186...)
- ZEN Leibn. Marg. 1 → 5 works (797894020, 79789604X, 797897755, 797898190...)

Method & caveats

- **Source:** OAI-PMH ListRecords (metadataPrefix=mets) over sets LeibnizHandschriften, LeibnizBriefwechsel, LeibnizMarginalien, Leibnitiana, leibniz-rekonstruktionen. Raw XML cached under data/oai/.
- **Page count = physical structMap page divisions.** Cross-checked equal to IIIF canvas count on sampled objects; a per-work re-check runs during leibniz harvest manifests and any drift is reported there.
- **Overlap is real, not double counting:** the same object appears under several sets; we dedup by object id. Marginalien are annotated *printed books* (hence high page counts), not autograph manuscripts — counted as page images all the same.
- **Scope:** Hannover GWLB holdings only. Non-Hannover Leibniz materials and transmitted-light fragment scans are out of Tier-1 scope (SPECS §1.1).

Image cache (Phase A2)

Generated 2026-09-11T15:01:49Z. Local delivery-derivative cache for the segmentation/HTR pipeline; images are never rehosted (the public viewer loads GWLB IIIF directly, SPECS §3/§4).

- **Cached:** 236,779 / 236,795 pages (100.0%), 395.6 GB on disk.
- **Mean page size:** 1.7 MB.
- **Projected full-corpus pull:** ≈ 395.7 GB for all 236,795 pages at the observed mean.
- **Dimensions (of 236,779 measured):** 529×202 ... 10593×15583 px.

Primary set	Cached	Target	%
Leibnitiana	1,768	1,768	100.0%
LeibnizBriefwechsel	72,284	72,284	100.0%
LeibnizHandschriften	58,812	58,828	100.0%
LeibnizMarginalien	103,887	103,887	100.0%
leibniz-rekonstruktionen	28	28	100.0%

Dimensions	Cached pages
< 1 MP	1,270
1–3 MP	15,835
3–5 MP	33,247
5–8 MP	38,437
8–12 MP	61,288
≥ 12 MP	86,702

Delivery mode of cached pages — iiif: 77,624, static: 159,155.