

A manifest, redacted.

This is the document that goes to a lab before any commercial conversation. The supplier here is anonymised: a working education technology company, seven fiscal years, roughly seventy contributors across the period.

MANIFEST, ANONYMISED SUPPLIER, AS OF 12 SEPTEMBER 2026		FDS-2026-0121
Mail, tenant wide		
VOLUME	YEARS	CONTRIBUTORS
30 mailboxes	2022 to 2026	27
DE-IDENTIFICATION		
Addresses tokenised, bodies scanned		
Chat, 46 channels		
VOLUME	YEARS	CONTRIBUTORS
7,900 messages	2022 to 2026	27
DE-IDENTIFICATION		
Handles tokenised, DMs excluded		
Application database		
VOLUME	YEARS	CONTRIBUTORS
140 tables	2021 to 2026	70
DE-IDENTIFICATION		
Identifier columns suppressed at source		

CRM

VOLUME	YEARS	CONTRIBUTORS
5,539 accounts	2022 to 2026	12
DE-IDENTIFICATION		
Buyer identity dropped		

Accounting

VOLUME	YEARS	CONTRIBUTORS
5 fiscal years	2021 to 2026	4
DE-IDENTIFICATION		
Bank and routing fields suppressed		

Document store

VOLUME	YEARS	CONTRIBUTORS
3,769 native files	2021 to 2026	70
DE-IDENTIFICATION		
Scanned pages redacted by pixel destruction		

Exact byte sizes per source are counted at export and stated in the manifest sent under NDA. They are withheld here because they identify the supplier.

What each column means

- **Source system.** The system of record the data was exported from, named specifically. Functional equivalents are fine, but we say which one it actually was.
- **Volume.** Counted from the source system at export, in the unit that system uses. Mailboxes, messages, tables, accounts, files, fiscal years.
- **Years.** The continuous period covered. A gap in the middle is stated, never smoothed over.

- **Contributors.** Distinct people whose work appears in that system during the period, whether on payroll or contracted.
- **De-identification.** The technique applied to that source before anything leaves our systems, drawn from the published standard.

What the manifest also carries under NDA

- Exact size per source, in bytes.
- The coherence map: which systems overlap in period and in people, which is what makes the corpus more than a pile.
- The full exclusion schedule for that supplier.
- Format distribution by file type and by industry segment.
- The de-identification techniques applied per field class, and the residual-risk note.

[Request the full manifest](#)