Archival Digitisation at Scale

Lessons from three years of the manuscript scanning project

Yelena Kovalenko

Institute for Historical Records

August 27, 2026

Outline

Scope

Throughput

Findings

What we set out to digitise

- 40,000 manuscript pages across five collections
- Fragile bindings requiring non-destructive scanning rigs
- Metadata linking each page to its original catalogue entry

Pages scanned per quarter

Year	Q1–Q2	Q3–Q4	Total
Year 1	2,100	3,400	5,500
Year 2	5,800	6,200	12,000
Year 3	8,900	9,100	18,000

Throughput more than tripled once the second scanning rig came online in Year 2.

Two scanning rigs compared

Overhead rig

- Faster per page
- Needs a flat cradle

Book-edge rig

- Handles fragile bindings
- Slower, more manual handling

What surprised us

1. Metadata entry, not scanning, was the actual bottleneck
2. Crowdsourced transcription cut review time by half
3. Colour calibration drift required weekly recalibration, not just at setup

Thank you

Questions?