

Transcription of 1600–1800s Handwritten Ledgers: AI Software, Vendor Process, and Lessons Learned

The Accessibility Challenge of Historical Documents

For higher education institutions, providing students with accessible versions of historical handwritten materials is a significant challenge. My unit experimented with several approaches to transcribe a large collection of ledgers from the 1600s–1800s. We developed a process to determine when to use internal resources, AI-powered software, or an outside vendor, and how to manage the vendor relationship to produce a high-quality, accessible final product.

The Challenge with Archaic Handwriting

Transcribing historical handwritten text has traditionally been an incredibly time-consuming process, creating a bottleneck for producing accessible materials.

- **Standard OCR is Insufficient:** While Optical Character Recognition (OCR) has made huge leaps for modern printed text, it consistently fails on archaic handwriting, especially the varied and often ornate cursive of this period.
- **Poor Source Quality:** Many historical documents exist only as poor-quality photographs or scans, which further complicates digital recognition.
- **Manual Transcription is Not Scalable:** The traditional method of manual reading and typing by staff is not viable for large projects, especially for the small teams typical in university accessibility departments.

Exploring AI Solutions: The Case of Transkribus

To address these challenges, we turned to new AI-powered tools. **Transkribus**, a platform developed by the University of Innsbruck, Austria, has created models that do a remarkable job of deciphering historical handwriting. However, we discovered several limitations.

- **Strengths:** The underlying AI models are powerful and can successfully transcribe text that is otherwise illegible to standard OCR.
- **Limitations:**

- **Difficult Interface:** The Transkribus user interface is entirely web-based and is riddled with bugs and operational inconsistencies.
- **Proofing is Essential:** As with any automated system, a manual proofing stage is non-negotiable to ensure accuracy.
- **Cumbersome Proofing Tools:** The platform's built-in proofing capabilities are difficult and inefficient to use.
- **Expertise is Still Required:** The quality of cursive varies wildly from writer to writer. Furthermore, understanding and proofing archaic dialects requires specialized knowledge. We found that while we could manage clear handwriting, expert intervention was necessary for more complex documents.

Developing a Vendor Process

Given the scale of our project and the limitations of the AI tool, we contracted with **SBL**, an India-based company specializing in transcription. Their process offered a promising hybrid approach:

1. Initial transcription using a combination of standard OCR and AI systems like Transkribus.
2. Manual transcription and proofreading by handwriting experts to correct errors and handle complex sections.
3. Delivery of the text in an accessible Microsoft Word document.

Lessons Learned from Vendor Collaboration

While the vendor provided quality transcription, we quickly learned that their understanding of accessibility standards did not fully align with our specific needs for student accommodations. Providing comprehensive accessibility training over long distances was not practical.

After reviewing the initial deliverables, our team reprocessed the documents to meet our standards. This revealed the precise limitations and understandings of the vendor and what our needs actually were. We then codified a precise set of parameters and formatting rules.

Our key takeaway is this: AI tools can be useful on small scales for this type of project when the work is not too complex. They are an improvement over old systems and may continue to improve in the future, but experts are still needed. Additionally, final, minor adjustments will likely always be necessary. Providing a vendor with a highly detailed set of formatting and structural requirements from the outset makes this final stage significantly faster and easier.

Recommendations for Vendor-Based Transcription

1. **Define Your Final Format First:** Before engaging a vendor, determine the exact accessible format you need.
2. **Provide a Detailed Style Guide:** Create a document for the vendor outlining how to handle page numbers, headings, tables, marginalia, and other unique structural elements of the source material.
3. **Specify Deliverables Clearly:** Ask for a simple, clean document (e.g., MS Word) with minimal formatting, which makes your final accessibility reprocessing much simpler.

Resources

For our complete decision tree, the final format specifications, and the process guide we developed for our vendor, please visit the following link to our [Vendor Determination and Process Guide](#) or scan the QR code below.

