The Positive Feedback Loop of Collaborative Infrastructure in Scholarly Research: **Transkribus®** as a Research Hub

Stauder, Andy (1); Rockenberger, Annika (2); Kaukonen, Minna (3); Ólafsson, Bragi Þorgrímur (4); Ingvarsson, Unnar (5); Foldvik, Therese (6)
Organisation: 1: READ-COOP SCE; 2: University of Oslo Library; 3: National Library of Finland; 4: National and University Library of Iceland; 5: National Archives of Iceland; 6: University of Oslo

CONTACT: a.stauder@transkribus.org
www.transkribus.org

Motivation: Investigating the synergy between shared infrastructures such as Transkribus and scholarly research, this study explores how such platforms catalyze a positive feedback loop in collaboration. This loop enhances the utility of collections, propelling interest in scanning projects across museums, libraries, and archives, which in turn boosts infrastructure use and scholarly quality through increased study of original sources and replicability.

Data Sharing
Data sharing is pivotal in scholarly collaboration, facilitating new insights using shared datasets. This essential process magnifies as more data becomes available, escalating the volume and quality of research outputs. An example of this is the training of AI models for HTR and NLP.
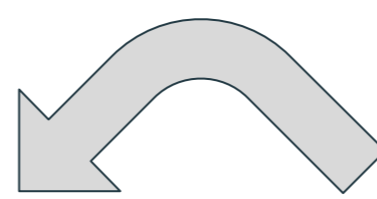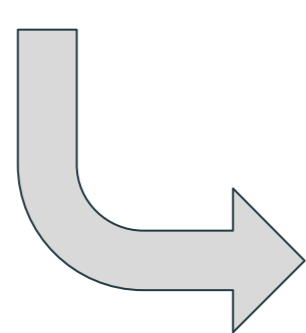
Compounding Effect
The collaborative cycle's compounding effect results in exponential knowledge growth. As collaborations expand, they accelerate data sharing, AI model training, research activities, and, consequently, knowledge production, manifesting in a rapid increase in global scholarly output.

The positive feedback loop engendered by collaborative scholarly infrastructure is a formidable driver of knowledge advancement. By fostering data sharing, direct collaboration, and educational initiatives, this cycle not only enhances scholarly research but also promises a future of ever-greater discoveries and innovations.
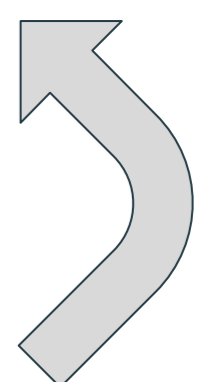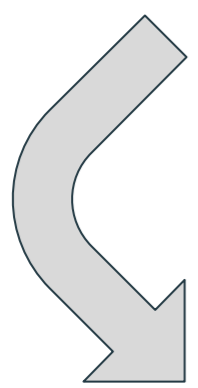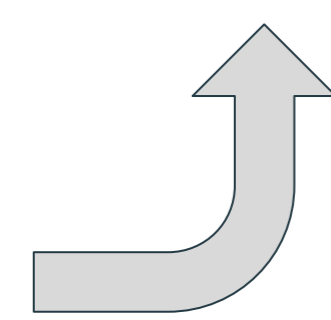
Direct Collaboration
Transkribus supports direct collaboration through its cloud-based platform, enabling diverse projects such as transcription, annotation, and digital publication. This collaborative model not only fosters community among researchers but also accelerates the creation of innovative methodologies and theories.

Human-Artificial-Intelligence Positive Reinforcement
The integration of human and artificial intelligence forms a powerful alliance that augments the efficiency and adaptability of research processes. This union not only processes larger quantities of data but also drives significant qualitative advancements in historical research.

Education
Education, especially in collaborative tools and infrastructure, is crucial in nurturing a future cadre of scholars. By integrating educational initiatives into research infrastructure, we prepare students and early-career researchers to engage in and perpetuate the collaborative cycle, enhancing scholarly endeavors.

Use Cases

University of Oslo Library

Ethica Complementoria and Tranchierbuch Transcription: Utilizing Transkribus, the library transcribed two Early Modern prints for a bilingual digital scholarly edition, minimizing manual corrections by re-using public models and training a dedicated model for German print. This project exemplifies how advanced text recognition supports the preservation and accessibility of historical texts.
Christopher Hansteen's Correspondence: The library collaborated with the National Library to transcribe the private correspondence of Christopher Hansteen, using a customized model for Hansteen's handwriting. This project not only aids in historical research but also in the digitization efforts at the Museum for University History and the History of Science.
East Asian Special Collection: Plans are underway to transcribe Tibetan prints from the 18th and 19th centuries, demonstrating the library's commitment to broadening access to specialized collections through digital technology.

National Library of Finland

NewsEye Project: Part of the Horizon 2020 programme, this project focused on reprocessing historically significant newspapers using Transkribus, which significantly improved search capabilities for research purposes. Post-project, the library continued to enhance the accessibility of Finnish and Swedish newspapers, amounting to a reprocessed corpus of approximately 2.5 million pages.
Ongoing Collaborations: Regularly collaborating with researchers and providing data services, the National Library of Finland exemplifies a successful model of integrating digital source materials and advanced text recognition tools into the research infrastructure.

National and University Library of Iceland

Icelandic Handwriting Recognition: This digital humanities project, aimed at developing Transkribus models to read old Icelandic handwriting, has resulted in publicly available models for late 18th century and 19th century scripts. This initiative not only preserves cultural heritage but also makes it accessible to a broader audience.
Collaborative Development and Access: Working through the Icelandic Centre for Digital Humanities and Arts, this collaboration promotes the development of and access to digital databases, linking Icelandic research with international advancements in digital humanities.

University of Oslo

Digitizing Norwegian Cultural Historical Material: SAMLA has undertaken the digitization of Norwegian archives containing cultural historical materials such as folktales and legends. The goal is to make this material accessible through a unified web-based portal, enhancing public engagement and educational outreach.
Research and Collaboration Framework: This project illustrates the synergy between research institutions and cultural organizations, aiming to create a robust methodological framework that bridges scholarly research and public accessibility.