iConTxt

Artificial Intelligence for Enhanced Access to Literature in Translation

White Paper

2026

Table of Contents

Executive Summary

The iConTxt project brings a transformative innovation to the **Digital Library and Bibliography for Literature in Translation and Adaptation (DLBT)** by integrating advanced artificial intelligence to unlock the full potential of its unique multilingual collections. As one of the most comprehensive and authoritative resources on literature in translation, the DLBT holds vast scholarly value, yet much of this material remains underused due to language and accessibility barriers. **iConTxt** directly addresses this challenge.

Developed collaboratively by eight European universities* in partnership with the Vienna University Library, the project leverages Large Language Models (LLMs) to automate metadata curation, generate English-language versions and summaries of reception materials, and produce new contextual knowledge derived from existing database materials. In this way, iConTxt transforms isolated records into interconnected, research-ready information.

Automated metadata curation

English-language summaries of reception materials

New contextual insights from existing databases

*University of Wrocław, University of Belgrade, Comenius University in Bratislava, University of Bucharest, Eötvös Loránd University (Budapest), Palacký University Olomouc, University of Salamanca and Stockholm University.

By embedding AI capabilities directly into the **YARM software** that powers the DLBT, iConTxt transforms the library into a dynamic, self-enriching research infrastructure. It enables user-friendly access for researchers, students, translators, literary agents, cultural institutions, and policymakers, fostering new opportunities for cross-cultural analysis, educational innovation, and literary exchange.

Its open-source design ensures that the resulting tools and methodologies remain freely available to registered users. It enhances global accessibility to cultural heritage, makes the trajectories of translated literature more visible, and contributes to a broader understanding of literature as a living, transnational ecosystem.

Who?

reserachers

students

translators

literary agents

cultural institutions

policy makers

What?

cross-cultural analysis

educational inovation

literary exchange

About DLBT

The Digital Library and Bibliography for Literature in Translation and Adaptation (DLBT) was developed at the University of Vienna as an open digital infrastructure for documenting the international circulation of literatures through translation, adaptation, and reception. It brings together two interconnected components: a **digital bibliography** and a **digital library**, each containing extensive records of translations, adaptations, and reception documents.

70,000+ entries

25,000+ digitized materials

ranging from newspaper articles and reviews to paratexts, images, and audiovisual content

Besides serving as a bibliographical repository, the DLBT also functions as a Digital Humanities workspace that integrates tools for visualisation and text analysis, such as Voyant Tools, which facilitate the exploration of translation and reception data through interactive visualizations and network maps. These tools support researchers in identifying patterns, trends, and connections in the transnational movement of texts, offering new perspectives on how literature travels, transforms, and interacts with other literary systems.

The DLBT is complemented by the **Dutch Literature in Translation (DLIT**) project, which focuses on mapping the diversity of the reception of Dutch-language literature abroad. Together, these projects have created an extensive, multilingual research environment that connects literary works, translators, publishers, and cultural mediators across Europe and beyond. The network's strength lies in the expertise of contributors from over a dozen countries – including Czech Republic, Poland, France, German-speaking countries, Hungary, Italy, Romania, Serbia, Slovakia, Spain, Sweden, Indonesia, and English-speaking states – who collaborate to expand and refine the bibliographical and digital corpus.

From DLBT to iConTxt

Despite its size and richness, much of the DLBT's multilingual reception material remains inaccessible to users who do not read the original language of the document, limiting international research use. In addition to this accessibility barrier, the DLBT's size and linguistic complexity have also begun to outgrow the limits of manual curation.

While the platform already integrates visualization and analytical tools to support data-driven exploration of translation and reception patterns, the next challenge was to automate and enrich the underlying metadata itself. The DLIT project had already demonstrated the scholarly value of connecting translation and reception data across borders; iConTxt now builds on that foundation by introducing artificial-intelligence-driven methods to ensure that these resources remain accurate, discoverable, and easily interpretable.

Through AI integration, iConTxt shifts from visualization to intelligent knowledge generation by connecting bibliographical data, contextual information, and multilingual content within a single adaptive ecosystem. This evolution marks a decisive step from mere digital archiving to active knowledge production within the field of international literary studies.

Methodology & Architecture

1 Data Import and Validation

Automated procedures verify bibliographical records during upload, identify inconsistencies, and detect duplicate entries. Integrated **OCR** correction modules improve the quality of digitized texts and facilitate full-text search.

2 Generation of iConTxtInfo Packages

In the first stage, iConTxt workflow leverages artificial intelligence models to process individual reception documents and generate structured data units designated as "**iConTxtInfo packages**". Each package contains an English-language summary, central question and issues, forming a unified knowledge base that improves accessibility and facilitates cross-document comparison. This database serves as the foundation for further AI-driven contextualization and research.

3 Metadata Enrichment

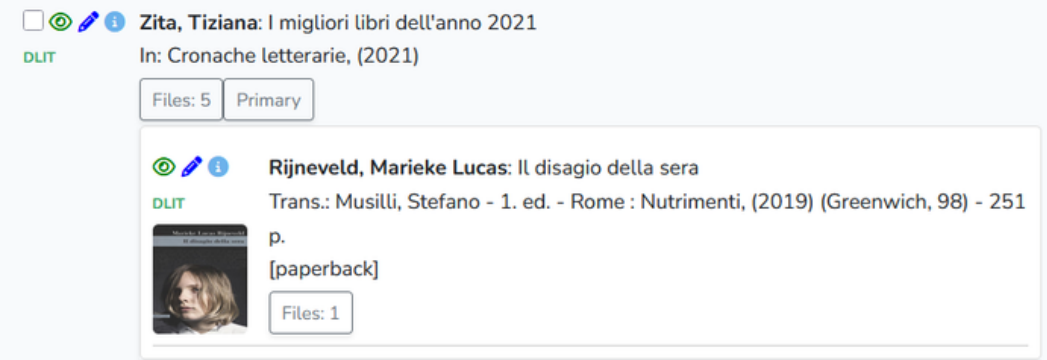
The system is used to refine and expand metadata by linking authors, translators, publishers, and works. It cross-references data with external authority files such as **VIAF**, **GND**, and Wikidata, ensuring semantic consistency and enabling multilingual interoperability.

4 Contextual Knowledge Generation

In the second stage, a **Retrieval-Augmented Generation (RAG)** pipeline combines the knowledge base with DLBT metadata, and external data sources such as selected web resources, Wikipedia, Wikidata, and specialized literary lexica, to generate new textual information. This approach ensures that generated information remains grounded in verifiable sources rather than purely model-generated inference. To ensure that newly digitized documents and metadata updates are continuously incorporated, these AI-generated texts will be regenerated every six months.

Example Use Case

To illustrate the workflow in practice, this example follows the processing of a reception document related to a translated Dutch literary work, namely Marieke Lucas Rijneveld's *De avond is ongemak* (*The Discomfort of Evening*).

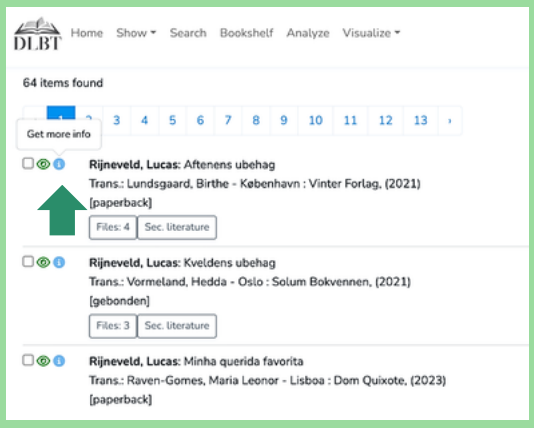


When the document is uploaded to the DLBT, the system first applies OCR correction to ensure text quality and detect key entities such as names, publication dates, and referenced works. In the first AI stage, the model automatically generates an iConTxtInfo package containing an English-language translation, summary, central research question, and key thematic descriptors. This structured information becomes part of the iConTxt knowledge base, linking the document to related entries such as the author, translator, and publisher.

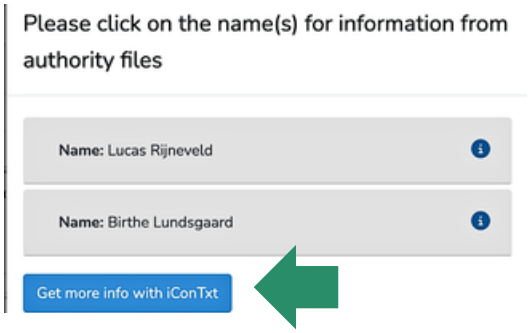
In the second stage, the RAG pipeline combines this internal knowledge base with external data sources to produce a new contextual text. This text provides a synthesized overview of Rijneveld's position within the Dutch literary field and her reception abroad, referencing relevant translations, publishers, and prizes.

This process demonstrates how iConTxt transforms individual reception documents into interlinked, machine-enriched knowledge units, making the complex networks of literary translation and adaptation visible and analyzable in entirely new ways.

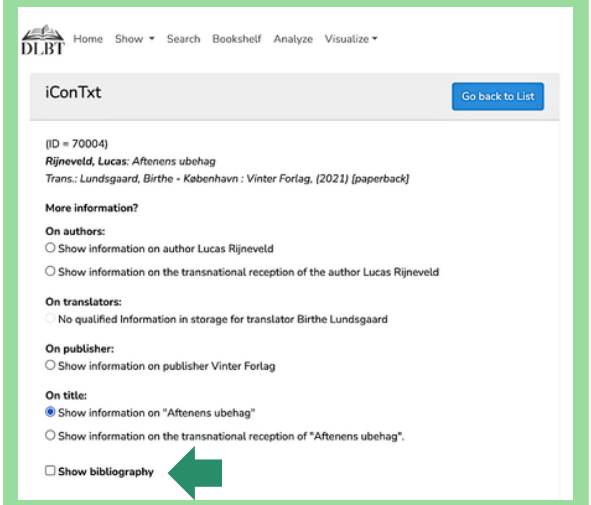
User Interaction and Access



Picture 1



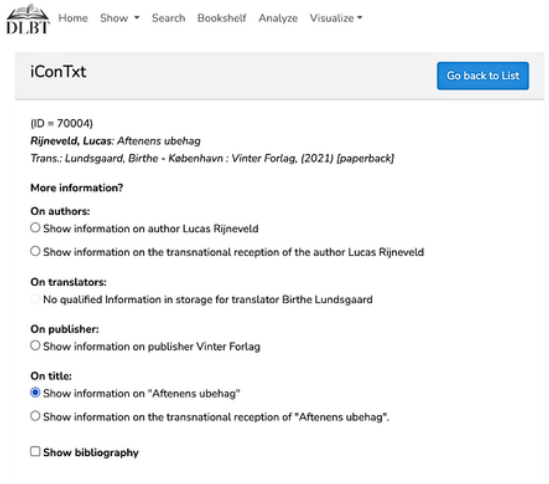
Picture 2



Picture 3

The AI-generated information (provided it is royalty-free) becomes accessible to the web users directly within the DLBT interface through an intuitive on-demand system. When consulting a record, users can retrieve additional contextual information via the following steps

1. Click “Get more information” on the record. (Picture 1)
2. In the pop-up window, select “Get more info with iConTxt”. (Picture 2)
3. Choose one of the available categories from the menu, for example:



Users may also select “Show bibliography” to display relevant translations. (Picture 3)

After clicking “Ask iConTxt”, the user is presented the AI-generated text.

Limitations and Challanges

While iConTxt introduces significant innovations in automating and enriching literary data, its implementation also exposes a number of conceptual, technical, and ethical challenges. Recognizing these limitations is essential for ensuring transparency, reliability, and long-term sustainability.

1

Linguistic and Corpus Imbalance

The DLBT encompasses reception materials in numerous languages, yet the system's generative processes currently prioritise documents identified as most relevant or information-rich. As a result, texts in low-resource languages or less represented cultural contexts may receive limited coverage in the generated outputs. This imbalance risks reinforcing existing asymmetries in the visibility of smaller literatures. Future development will focus on progressively including more diversity in corpus and improving automatic weighting mechanisms to reduce bias.

2

Copyright and Legal Constraints

A considerable proportion of the materials in the DLBT are subject to copyright protection, which poses a risk of unintentional infringement during processing and generation. In addition, copyright regulation is not harmonised internationally: national differences in fair-use or reproduction rights complicate uniform policy implementation. To mitigate these risks, the project will initially focus on authors whose works are no longer protected by copyright and will seek licences or permissions for more recent materials.

3 Quality and Reliability of Digital Sources

The quality of the texts used for AI processing, especially web-based materials, remains uneven. The internet is increasingly saturated with automatically generated or poorly curated content, and digital sources may disappear or change without notice. This volatility affects reproducibility and long-term validation. Continuous monitoring and metadata versioning are therefore essential to maintain the integrity of the knowledge base

4 Hallucinations and Stylistic Stereotyping

Large Language Models may occasionally produce factual inaccuracies or stylistic clichés. To mitigate these effects, the editorial team employs a multi-prompt strategy and a validation pipeline designed to reduce hallucinations and improve factual precision. However, certain issues are inherent to current LLMs, such as incorrect transcription of proper names, overly literal or inaccurate translations of book titles, and difficulties distinguishing between homonyms or similarly spelled authors. Such inaccuracies can lead to misattribution, reduced searchability, and sometimes also inconsistencies in bibliographic records.

5 Ecological Considerations

AI-based text generation entails substantial computational cost and energy consumption. To limit its ecological impact, iConTxt adopts a selective regeneration strategy: AI-generated texts are renewed periodically to integrate newly digitised materials, but only for entries that demonstrate ongoing user activity. Texts that are not accessed or cited will have lower priority, thereby reducing unnecessary model use and preserving computational resources.

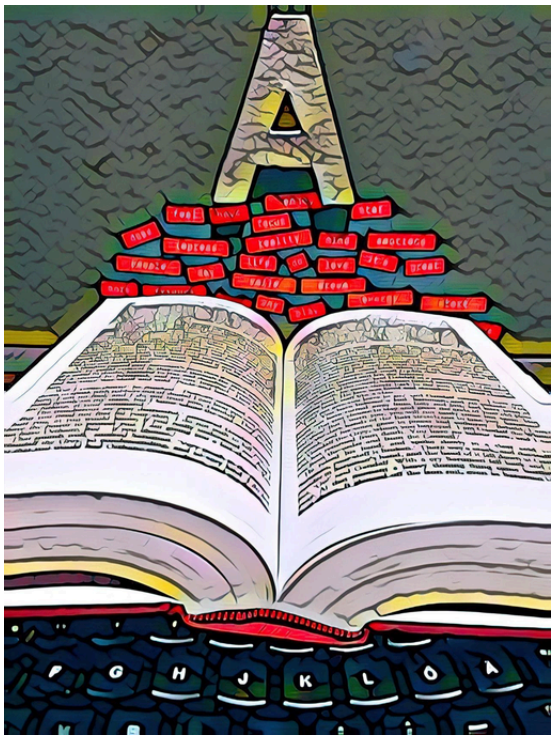
Expected Outcomes and Impact

By embedding AI capabilities directly into the DLBT ecosystem, iConTxt opens doors to literary discoveries that were previously closed. For the first time, researchers, students, and literature enthusiasts can explore the vast multilingual treasures of the DLBT, regardless of their technical expertise or the languages they speak. This platform transforms complex data into accessible insights, allowing anyone to trace how Dutch literature travels the world, influences cultures, and sparks conversations across borders. iConTxt delivers benefits across multiple dimensions:

Scholarly and Research Impact

For researchers, iConTxt unlocks new opportunities to trace translation flows, reception patterns, and transnational literary interactions. By providing English translations and summaries of multilingual reception materials, the platform enables scholars to analyse documents that were previously inaccessible due to language barriers. The platform enables comprehensive analysis of the entire digital libraries within the DLBT, enhanced through automated import tools that integrate data from other sources. Researchers can now:

- Conduct systematic analyses of studies and reviews about Dutch literature in translation abroad
- Investigate the transnational role of authors, books and publishers across different cultural contexts
- Track reception patterns across different countries and time periods
- Compare different translations



Teresa Berndtsson /
https://betterimagesofai.org
https://creativecommons.org/licenses/by/4.0/

Educational and AI-Literacy Development

Beyond serving researchers, iConTxt becomes a practical training ground for digital and AI literacy within translation and literature programs. The project provides a unique environment where students can:

- Practice evaluating AI-generated content and learn to identify machine-generated texts
- Critically assess the accuracy and quality of automated translations and summaries
- Understand transnational literary reception through hands-on exploration

These AI-skills modules prepare students for a future where artificial intelligence tools are integral to literary and cultural research.

In cooperation with the PETRA-NED network, a dedicated set of educational modules has been developed to support this training function. The modules are designed for university courses and workshops on literary translation from Dutch and are aligned with the PETRA-E European competence framework.

Societal and Public Engagement Impact

iConTxt transforms complex literary reception data into actionable insights for professionals working in the book industry. By aggregating and analysing information on authors, translations, publishers, and international reception, the platform supports informed decision-making for rights acquisition, title selection, pitching, and market positioning.

For publishers, scouts, literary agents, and translators, iConTxt offers:

- Strategic insight into international potential – AI-generated overviews reveal how authors and titles are received across different language areas, helping assess whether a work has momentum, niche potential, or growing international relevance.
- Cross-market comparison tools – visualize where a book has been translated, reviewed, awarded, or adapted, enabling comparison of success across regions and identifying gaps or emerging markets.
- Faster and more informed decisions for publishing houses – access reception summaries, key themes, and positioning arguments that can directly inform rights negotiations, pitch materials, and catalogue curation.
- Evidence-based pitching support – translators and agents can use ready-made contextual information to prepare stronger pitches to publishers, cultural funds, and international partners.
- Reduced language barriers – English-language summaries and reception overviews allow professionals to evaluate foreign-language titles without needing full proficiency in the original review language.

By providing reliable, multilingual, and data-driven insights, iConTxt empowers professionals to spot trends earlier, select titles with greater confidence, and communicate the value of literature more effectively on the international stage.

Conclusion and Future Direction

The iConTxt project represents a crucial step in transforming the DLBT from a static bibliographical repository into a dynamic, intelligent research infrastructure. By combining large-scale data management, natural-language processing, and human editorial expertise, iConTxt demonstrates how artificial intelligence can enhance but not replace, scholarly interpretation within the humanities.

In its next development phases, the project will continue to expand both technologically and conceptually. The immediate priorities include:

- extending the corpus to incorporate additional languages and underrepresented regions, thereby addressing linguistic imbalance and ensuring more inclusive coverage of global literary reception;
- refining the multi-prompt strategy and editorial validation workflows to further improve factual accuracy and stylistic quality;
- deepening interoperability with other research infrastructures, such as Digital Library for Dutch Literature (DBNL), National Library of the Netherlands (KB), and Dutch and Flemish Literary Foundations, to ensure seamless metadata exchange; and
- developing user-oriented interfaces, including chat-based research assistants, that will allow scholars, students, and general readers to interact with the DLBT's growing knowledge base in an intuitive, conversational way.

Longer term, iConTxt aims to serve as a model for sustainable and ethically grounded AI use in the digital humanities. By coupling automated knowledge generation with cyclical human verification, transparent methodologies, and ecological restraint, the project offers a replicable framework for multiple literary and cultural databases. Through this integration of technology, scholarship, and ethical reflection, iConTxt contributes to a broader understanding of literature as a living, transnational ecosystem that continues to evolve across languages, media, and cultures.

Glossary of Technical Terms

GND (Gemeinsame Normdatei): The Integrated Authority File used by German-speaking libraries for standardized cataloging of persons, corporate bodies, conferences, geographic entities, topics, and works.

OCR (Optical Character Recognition): Technology that converts images of text (from scanned documents or photos) into machine-readable and editable text data.

PHAIDRA: An acronym for Permanent Hosting, Archiving and Indexing of Digital Resources and Assets. It is a management system for permanent preservation of scholarly objects and associated metadata.

RAG (Retrieval-Augmented Generation): An AI technique that combines information retrieval with text generation, allowing AI models to access external knowledge bases before generating responses, improving accuracy and reducing hallucinations.

VIAF (Virtual International Authority File): An international authority file that combines multiple national library authority files into a single service.

VOYANT: A web-based text analysis and visualization tool that helps researchers explore patterns in digital texts through various analytical views.

YARM – Reference management system developed at the Vienna University Library. YARM provides the database and the tools of the DLBT and links the DLBT with PHAIDRA.

Impressum

PROJECT LEADERS

Herbert Van Uffelen

Universität Wien

dlbt@univie.ac.at

Małgorzata Dowlaszewicz

Uniwersytet Wrocławski

WEBSITE

https://dlbt.phaidra.org

WHITE PAPER

Bojana Budimir

University of Belgrade

This project and the present White Paper were realized with the financial support of the Dutch Language Union (Taalunie).