

# NATIONAL LABORATORY FOR DIGITAL HERITAGE

## ENCODING OUR HERITAGE FOR INNOVATION

The aim of the Laboratory is to promote the processing, research usability and broad smart accessibility of digital cultural heritage using digital technologies, especially artificial intelligence, in Hungary and in Hungarian speaking communities. The Laboratory's main task is also to organise a web-harvest with the aim of collecting web resources relevant to research and innovation in the Hungarian community.



## MAIN RESEARCH AREAS

- Gold standard corpus
- Text mining
- Annotated existing corpora
- Web-harvest of Hungarian texts
- "Born digital" labs
- Making digital cultural heritage searchable
- Natural language processing
- Processing literary corpora using artificial intelligence
- Digitisation and processing of Hungarian written heritage beyond the borders
- Handwriting recognition and character recognition systems



### CONSORTIUM LEADER:

Eötvös Loránd University

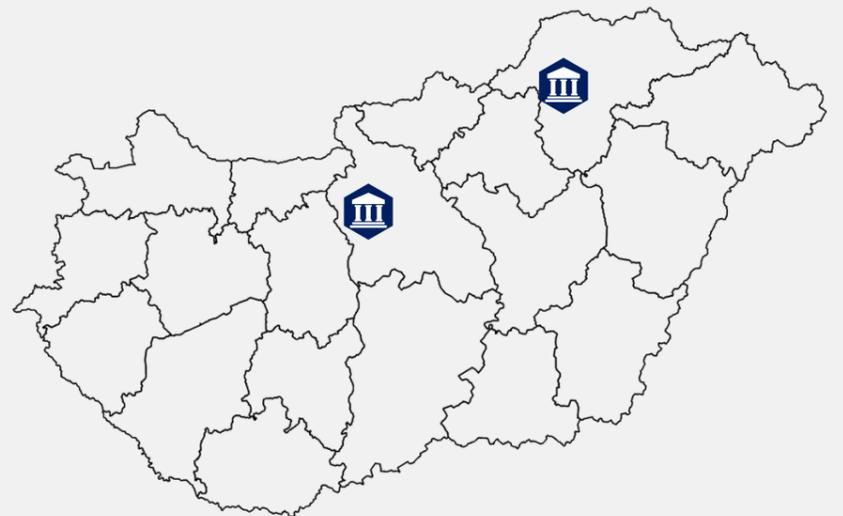
### CONSORTIUM PARTNERS:

University of Miskolc

Research Centre for the Humanities

National Archives of Hungary

### PLACES OF IMPLEMENTATION: Budapest, Miskolc



## BENEFITS TO BE EXPECTED FROM LABORATORY RESEARCH

- Expand and modernise the manually compiled (gold standard) text corpus for machine learning and process control.
- Organise the web-harvest of the collection and semantic machine processing of research and innovation relevant web resources of the Hungarian community in Hungary and beyond the borders of Hungary, and their searchability by means of artificial intelligence.
- Harvest Hungarian texts from the web.
- Increase Hungarian presence in the world of digital cultural heritage.
- Facilitate the searchability of digital cultural heritage and its widespread intelligent accessibility through artificial intelligence.
- Radically increase the proportion of Hungarian-language materials to promote the international presence of Hungarian cultural heritage and research potential.

## THE PROFESSIONAL TEAM

**Gábor Palkó (PhD)** Researcher at the Research Centre for the Humanities and Co-Head of Department of Digital Humanities at the Eötvös Loránd University. He majored in Hungarian Language and Literature receiving his PhD degree in 2002. He teaches digital cultural heritage and digital humanities at the Eötvös University. His wider field of interest includes digital philology and cultural heritage, theories of the archives, Niklas Luhmann's media and art theory and 20th-century Hungarian literary history. He is the Editor-in-Chief of the International Journal of Digital Humanities published by Springer. He took part in several international projects as project manager/coordinator and researcher (Athena, Athena Plus, Europeana Inside, DM2E), and recently: Europeana Common Culture. He is the editor-in-chief of the service DigiPhil which is an aggregator of Europeana.

**Balázs Indig (PhD)** NLP expert, research fellow at Eötvös Loránd University Centre for Digital Humanities and senior technology developer at Research Institute for Linguistics since 2018. He has been working on Hungarian NLP projects as part of his MSc studies at Pázmány Péter Catholic University (PPKE) since 2011. In 2018 he completed his PhD on Psycholinguistically motivated NLP pipelines at PPKE. In 2013 he has won the 2nd prize of "National Conference of Scientific Students' Associations" in the computer science section. In 2016 been awarded with the "National Young Talent Scholarship (part of the National Talent Programme)" founded by the Ministry of Human Capacities. He co-authored more than 50 publications and numerous open-source NLP tools since 2011. He specialised himself in in linguistics-centric NLP tools, pipelines and corpus creation. Since 2018 he is the lead developer of e-magyar language processing system. Currently he is working on a new Hungarian webcorpus created with webarchiving technics.

## POSSIBLE PARTNERSHIPS

- International cooperation in the field of digital cultural heritage and digital research (eScience)
- International centres developing multilingual tools: to jointly improve the efficiency of Hungarian language analysers
- International cooperation network: web research-web archiving, archiving and making born digital material searchable, Web 3.0, or semantic web

## NETWORKING

- DARIAH: Digital Research Infrastructure for the Arts and Humanities
- Distant Reading for European Literary History COST Action
- centerNET: An international network of digital humanities centers
- MCA: Michael Culture Association
- Archives Portal Europe Foundation
- DLM Forum: Document Lifecycle Management Forum Foundation
- ICA: International Council of Archives
- TMO: Time Machine Organisation
- ICARUS International Centre of Archival Research

## TARGET GROUP

- Researchers dealing with digital humanities, philology, linguistics
- The general public

## PROFESSIONAL CONTACT

**DR. GÁBOR PALKÓ**  
*Project Coordinator*



contact@dh-lab.hu



+36 1 411 6500 / 5088