

eite.dh

Centre for Digital Humanities

Contents

What is ELTE.DH?

International relations

- DARIAH
- DARIAH CEH
- centerNET
- Course Registry
- MCA

Research

- Repository development
- New National Excellence Program

Institutional Excellence Program for the Higher Education

- Natural Language Processing
- Handwritten Text Recognition
- Annotation of big scale poetical corpora
- Stylometry
- Prosopographical database development

Education

- DH workshops
- University courses
- Training

DH colloquia

- Guest lectures
- ELTE lectures in the Spring Semester
- DH_Budapest_2018 Conference
- Workshops, presentations

Publications

- IJDH
- Digital (Text) Cultures in the Humanities

Grants

- COST Action: Distant reading
- Visegrad Fund

Partners

Members



What is ELTE.DH?

Introduction

It was among the most important tasks of Hungary's oldest institute of higher education, the Eötvös Loránd University, to adopt the latest scientific and technical advancements in the field of humanities.

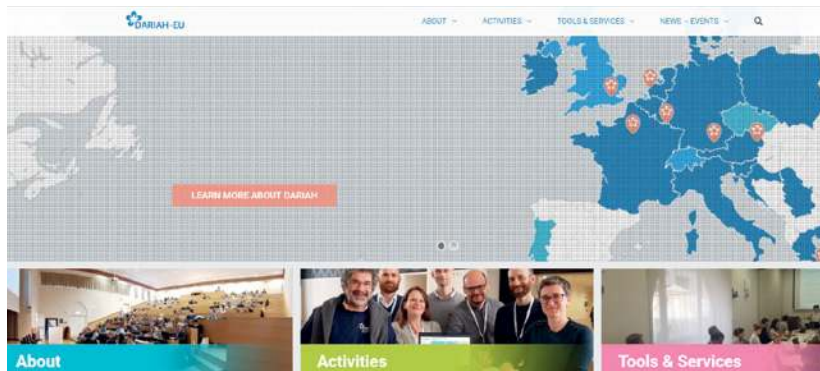
The practice of digital humanities – which now has become a worldwide phenomenon – unites humanities with a specialized form of computer science. Furthermore, there are international networks of centres, conferences, and summer universities dedicated to digital humanities, journals published online and in paper form as well.

It was highly important for the Faculty of Humanities at Eötvös Loránd University to keep up with the global trends and to create a stable centre for the education and development of digital humanities, therefore the University formed the Centre for Digital Humanities in 2017.

As a research- and practice-oriented scholarly institution, ELTE.DH ensures focus and coherence for all efforts that have been made in the field of digital humanities at the faculty, and examines digital research aspects not yet considered.

International Relations

DARIAH On the initiative of the Centre, ELTE has become a cooperative partner of the world's largest network of digital humanities, DARIAH. DARIAH is a member of the European Research Infrastructure Consortium, a Pan-European infrastructure for arts and humanities scholars working with computational methods. It supports digital research as well as the teaching of digital research methods.



DARIAH CEH With the representatives of the digital humanities organizations of Austria, the Czech Republic, Poland and Slovakia, ELTE.DH established the DARIAH Central European Hub with the aim of developing regional cooperation and joint tender activities. The Centre with the support of the DARIAH Humanities at Scale project invited representatives of leading institutions in digital humanities from Austria, the Czech Republic, Poland, Slovakia and Hungary to discuss institutional systems, beneficial practices, partnership opportunities across borders and the advantages of an international infrastructure. The principal aim of the hub is to foster collaboration among institutions and individual scholars interested in digital technology in the field of humanities. Participants especially stressed the importance of possible joint activities focusing on three key collaboration issues in Central Europe: establishing a knowledge base of digital humanities research, aligning teaching activities by sharing knowledge and developing or using common eLearning materials, and fostering the usage of multilingual computational language tools in the humanities and the social sciences. The first result of this is that ELTE.DH jointly applied for organising summer universities at the Visegrad Fund.



Specialists who took part in the founding meeting and/or took active part in organising it:

Silvie Cinková

(Charles University, Czech Republic)

Andrej Gogora

(Slovak Academy of Sciences, Slovak Republic)

Jan Hajič

(Charles University, Czech Republic)

László Kovács

(Hungarian Academy of Sciences Institute for
Computer Science and Control, Hungary)

Katalin Lejtovicz

(Austrian Centre for Digital Humanities, Austria)

Martin Lhoták

(Library of the Czech Academy of Sciences, Czech
Republic)

Maciej Maryl

(Institute of Literary Research of the Polish Academy
of Sciences, Poland)

Gábor Palkó

(Centre for Digital Humanities, Eötvös Loránd
University, Hungary)

Gerlinde Schneider

(University of Graz, Austria)

Tanja Wissik

(Austrian Centre for Digital Humanities, Austria)

Marcin Werla

(Poznań Supercomputing and Networking Center,
Poland)

Miroslav Zumrík

(Slovak Academy of Sciences, Slovak Republic).

centerNet

centerNET ELTE.DH as the only organization from Hungary, has become a member of the centerNet network, which connects roughly 250 DH organizations around the world. CenterNet is an international network of digital humanities centres that provides collaborative actions for the overall benefit of digital humanities and related disciplines, moreover, it serves as a cyber infrastructure for humanities. CenterNet enables individual DH Centres to form an international network – to share projects, tools, employees and expertise. Through initiatives such as Day(s) of DH and Resources for Starting and Sustaining DH Centres, centerNet provides a virtual DH center for isolated DH projects and platforms to educate the broader research community in digital humanities.



DIGITAL HUMANITIES COURSE REGISTRY

DH Course Registry Co-director of ELTE.DH, Gábor Palkó became responsible for the DH Course

Registry in Hungary: ELTE.DH started the registration of the digital humanities-related university courses in the database. Previously, no Hungarian courses were included in the service. The DH Course Registry encourages enrollment in DH classes and showcases DH courses. Recently the existing platform for the sharing of information on DH courses has been relaunched as a joint effort of the European research infrastructures CLARIN ERIC and DARIAH-EU. The Course Registry currently contains a selection of DH courses offered by European academic organizations.



Michael Culture Association

Michael Culture Association ELTE.DH has become a member of the Michael Culture Association, which

represents the digital cultural heritage system in Europe. Michael Culture Association is not-for-profit organisation, founded in 2007. It has become a key actor in the promotion and valorization of the digital cultural content, and gathers a strong network of more than 100 institutions all over Europe.



Research

Repository development

In cooperation with the ELTE University Library, ELTE.DH began to develop a semantic-based data repository for Digital Humanities data curation and publication, in which the Online Knowledge Base of Hungarian Philosophy is being created as a pilot project. This project aims to collect texts, sound- and video recordings that are relevant to the history of Hungarian philosophy. The materials of the repository will fill gaps in this scientific area, and will hopefully help to make an outline for the history of the Hungarian philosophical canon.

The software base for this and other projects in the future is the open source Islandora 8 which was released recently and had been greatly improved compared to previous versions. The most essential upgrade was the change in the infrastructure. Islandora 8 consists of a Drupal 8 content management system, a Fedora 5 database, a Blazegraph database, a Solr search engine and a synchronisation component.

Some of the features provided by Islandora 8 are: site management, localization, the creation of personalised forms by using unique or standard metadata descriptions, like dublin core or the schema.org along with Fedora RDF mapping. Certain features can only be accessed by downloading external modules. These include for example displaying dynamic content and audiovisual materials, batch import, automatic indexing, etc.

Repository development at the Centre for Digital Humanities (ELTE.DH)



Gábor Palkó – Zoltán Kanász-Nagy – Anikó Mohay – Tünde Molnár
Eötvös Loránd University
Faculty of Humanities



The project

The Centre for Digital Humanities at Eötvös Loránd University with the help of Qulto is creating a new Islandora CLAW based online repository in order to store and preserve content according to the most recent technological developments.

Hungarian Philosophical Database (Magyar Filozófiai Tudástár)

The Hungarian Philosophical Database aims to collect texts, sound- and video recordings that are relevant to the history of Hungarian philosophy. These materials will fill certain gaps in this scientific area, and hopefully help to make an outline for the history of the Hungarian philosophical canon.



Islandora CLAW

The software base for this and other projects in the future is the Islandora CLAW which is an open-source software framework designed to help institutions and organizations. It is a new, beta version that had been greatly improved compared to previous versions.



Key components:

- Fedora 4
- Drupal 7 which includes:
 - A relational database such as MySQL or PostgreSQL
 - Several third-party Drupal modules from drupal.org
 - Several Islandora Drupal modules
- Apache Solr
- Blazegraph triple store
- Apache Camel based middleware

Fedora 3 vs. Fedora 4

In the most basic terms, Islandora CLAW is the version of Islandora that works with Fedora 4. Because Fedora 4 is a vastly different platform than Fedora 3, so too is Islandora CLAW a major departure from what came before. Switching to CLAW represents not just a typical upgrade with improvements, features, and bug fixes, but rather a major shift in how objects are stored and managed.

Fedora 3	Fedora 4
Inline and managed XML	RDF properties
Objects stored in flat directories	Objects stored in a hierarchical filesystem
Objects have Persistent Identifiers (PIDs)	Objects have a path based on their location in the file system hierarchy
	Objects can also have other identifiers (DOIs, Handles, PIDs, etc.)

Functions

- User and website management
- Creating personalised forms
- Unique and standard metadata description (schema.org, Dublin Core)
- Creating Fedora RDF mapping
- Indexing Drupal content into Fedora and Blazegraph
- Solr search
- Localisation
- Batch processing
- Batch import (csv)
- Dynamic fields (Display Suite)



References

<https://islandora.ca/CLAW>
<https://islandora-claw.github.io/CLAW/user-documentation/CLAWfor1x/>

New National Excellence Program (2017-2018)

The aim of the Program initiated by the Ministry of Human Capacities is to support the Hungarian education, to stimulate and reinforce academic and artistic achievements; to assist grantees in start of their career path and also to help them to maintain their research or art.

Three researchers were approved with DH-related topics.

Melinda Vasari: Digital Forensics

The main interest of the project is the influence of the medial environment of our times to the practice of literary studies. The main questions of the research are, how literary studies, especially philology can adopt and make use of the new digital media environment. The motivation of the research is to reinterpret the already existing practices in the light of these new developments which means both their application in practice and theoretical reflection on them.

Dániel Kozák: Digital Corpora and the Practices of Classical Philology

The focus of the project is the effect that digital corpora have on intertextual interpretation. How does the methodology and practice of finding hitherto undocumented intertextual connections change if one is engaged in this interpretative practice (very typical of classical philology) using digital tools? Another aim is to map the opportunities for the Department of Latin and the Centre for Digital Humanities at ELTE BTK to start a new digital project or join existing ones.

Szilvia Maróthy: Digital Scholarly Editions in Hungary

The aim of the research is to collect and process Hungarian digital scholarly editions with the pretense of completeness. The catalogue, which uses the suitable international standards and is free to access, will be implemented into the Centre's repository connected to other services.



development to keep up with the current natural language processing (NLP) trends. These are claimed to be 'language independent', but in reality, they are 'English-driven'.

The Center relies heavily on these tools for many applications, including but not limited to stylometry and semantic web. The most current Hungarian NLP pipeline is developed by professional computational linguists using the well-founded mature methods that have proved to be functional for the Hungarian language, but still these tools have a small number of applications. As there are limited resources, our primary focus in their development, is the standardisation and code reuse to be able to utilize as much usable language independent parts from existing tools as possible, whilst keeping the highest quality that can be achieved in technical and linguistic terms.

We are participating in the application centric development and standardization of the e-magyar language processing system, and our goal is to make it a modular, open, and standard compliant platform that fulfills any digital humanities related needs and can be further extended on demand. The center also gathers current Hungarian language text from the internet into large corporas for web archiving purposes. To be able to process them with the e-magyar system into richly-annotated data, currently popular neural network based vector embedding models can be used, which are needed by the digital humanities and social sciences.

Handwritten Text Recognition Using Machine Intelligence

In cooperation with the Institute for Literary Studies of Hungarian Academy of Sciences (MTA ITI), ELTE.DH is currently working on the automatic recognition and transcription of handwritten letters by János Arany (1817-1882). The project relies on Transkribus, which is a comprehensive platform for the automated recognition, transcription and searching of historical documents. It uses machine learning, therefore in theory, it can be applied to documents in any language, layout, or style. After uploading the manuscript images to the Transkribus server, it can automatically recognise text regions, as well as the baseline in each row. For the automated text recognition, it requires at least 100 pages of transcribed images that are used as training data for the machine learning algorithms. Transkribus offers a number of tools not only for the word by word transcription,



but also for tagging metadata: personal names, places or locations, dates, as well as organizations, all that can be marked with the corresponding tags.

When the training data is ready, a Handwritten Text Recognition (HTR) model can be created with the help of the Transkribus team. Accuracy differs from project to project, as it depends on a number of different factors like the quantity of the training data, consistency of the handwriting, etc., but there are promising projects with a Character Error Rate of around 5%.

Annotation of Big Scale Poetical Corpora

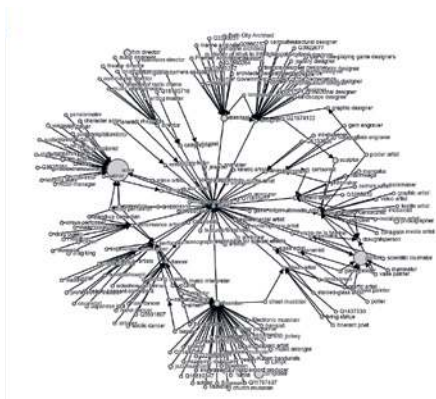
The goal of the research project is to create an annotated corpus of 19th and 20th century Hungarian poetry. The annotation is based on automatic methods. We annotate the structural elements of poems and we are developing a program that annotates rhythm, rhyme patterns, and alliterations. In case of words, we also annotate part of speech, morphological and phonological features. In collaboration with the Research Group in Stylistics at ELTE, we are developing a subcorpus in which the verbal constructions are annotated manually. This research focuses on person-marking constructions in poetic discourses. Currently, we are working on the annotational scheme for the manual annotation. In the future, we plan to broaden the scope of the project by creating subcorpora containing less canonical poetic texts, such as song lyrics and slam poetry texts. It is hoped that the corpus will be useful not only in literary and linguistic researches, but in education as well. One of our aims is to make the corpus accessible for everyone.

Stylometry

The goal of this research is to better understand the R programming language and the ,stylo' package in order to apply them in stylometric analysis. Stylometry is able to create data about the linguistic and thematic structure of texts that would not be possible for humans. In contrast to the human interpretation, the analysis made by a computer focuses on different levels of the text (e.g. the micro environment, as word frequency or the macro environment, as an analysis of even thousands of texts at once), thus creating different patterns. The R programming language and the ,stylo' package were created for this exact purpose. Plans for the future include the stylometric analysis of the works of Péter Nádas, for the size of the texts, the thematic complexity, and linguistic structure.

Prosopographical Database development

The research focuses on the 15-16th-century humanist network of East Central Europe and on the possible digital exploitation (network analysis, construction of a database, data visualization etc.) of the forthcoming book of the Lendület Research Team (HECE) led by Gábor Farkas Kiss, entitled Companion to Humanism in East Central Europe. The researchers analyse the biography of Central European humanists to find common features in their career and to establish a model for the construction of a prosopographical database.



Following our preliminary experience, they are currently preparing the HECEdata Policy, focusing on a consequent data curation. The SPARQL endpoint of HECEdata enables a particularly effective control on the relational network of the raw database.



Education

DH Workshops

The workshop series organized by ELTE.DH in the first three weeks of May 2018, served well as a prequel to the DH_ Budapest_2018 conference. In these open courses, experts provided introductions into their field of expertise combined with practice, representing in many ways the topics of Digital Humanities for PhD students and teachers.

On May 3, 2018, PhD student of Pázmány Péter Catholic University and expert of the Petőfi Literary Museum Zsófia Fellegi conducted the first session. She approached an area perhaps most familiarly interfacing with humanities among the three workshops, since she spoke about the theoretical and practical problems of digital scholarly editions.

According to her instructions, the participants prepared the XML coding of a poem by Sándor Petőfi (1823-1849), thus gaining an insight into the digital philologist's workroom through actual examples.

On 10 May, Zsolt Czinkos (Central Statistical Office) continued the series with a workshop on data visualization. This workshop illuminated the foundations

of visual perception, while advancing a deeper theoretical understanding of reading on the Web and of information adoption processes in general. Examples from the vocabulary of János Arany (1817-1882) gave participants the opportunity to test their own data visualization methods using the Flourish online application. Although output of this app is less customizable and its share is more limited than pay-per-view applications, it is spectacular and instructive for educational purposes and for creating diagrams that can be saved.

On May 17, the presenter was Tsegaye Misikir Tashu, who held an English-language workshop on text-mining. Using examples from American presidential speeches, he demonstrated the steps of preparing the texts, erasing fill words, and the method of deriving the infinitive forms from vocabularies worth for comparison. The steps could be implemented by the participants themselves based on the instructions of Tsegaye, with Python commands in the Jupyter Notebook pre-installed on the computers. By interpreting the code, the participants came closer to understanding text processing and could see that with the help of free libraries written for Python, texts that have statistically measurable size can be compared after a few steps. The expert also showed illustrative examples for evaluating and visualizing the results.

University courses

Gábor Palkó conducted preliminary courses (Introduction to Digital Philology) available for all ELTE students in two consecutive semesters. The aim of these courses is to introduce contemporary cultural practices of digital humanities. It presents several DH methods and tools relevant across humanistic disciplines: background and hands-on use of markup languages (SGML, HTML, XML, JSON, XSL), text analysis and encoding, digital philology, data and metadata, digital object, data exchange formats, aggregation, visualization and mapping.

Competencies:

- Ability to integrate digitally driven research goals, methods, and media with discipline-specific inquiry
- Ability to understand, analyze, and use data
- Develop critical savvy for assessing sources and data
- Ability to use design critically
- Ability to assess information and information technologies critically
- Ability to work collaboratively

Digital research infrastructure for digital philology: WYSIWYM interface for TEI XML encoding of a critical correspondence edition



FRAMEWORK

TEI

INTRODUCTION - **ABOUT THE PROJECT**

... (transcription of the text) ...

... (transcription of the text) ...

Manuscript-Linguistics Institute The Janssen Project

ABOUT THE PROJECT

... (transcription of the text) ...

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

TEI VISUALIZATION OF AN EXTENDED TAGSET (FULL TAG MODE)

... (transcription of the text) ...

Abstract

The main project of the *Fachbereich Literaturwissenschaft* and the *Manuscript-Linguistics Institute* for literary studies named *DigiPhil* (Digital Philology) is a knowledge base of scholarly text editions, bibliographies, and datasets in the framework of the *Manuscript-Linguistics Institute*. This environment provides a safe, easy-to-use and trustworthy user interface for every function of the digital philological workflow. This WYSIWYM (What You See is What You Mean) interface facilitates encoding and data enrichment, while providing software and hardware independence on the client side. At the same time, it produces synthetically correct codes on the server side in a secure data environment through functions as verification and automatic error correction (validation). The codes produced are also automatically convertible to standard TEI XML format and the environment also provides an easy and fast way to implement error correction in the proofreading process. The process is shown by presenting the digitization of the *Heavenly James Critical Edition*.

DigiPhil (Initially Text Editions, Bibliographies and Research Database Online)

Activities

- publishing markup language transcripts of scholarly text editions
- publishing writers' bibliographies
- aggregating metadata
- developing a semantic annotation tool

Virtual research environment

Why?

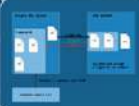
- there are very few philologists who want to learn the TEI encoding
- there is no available WYSIWYM (What You See is What You Mean) editor which is customizable for the needs of DigiPhil

What are the advantages of the framework?

- easy to install as part of the *OpenOffice* editor
- way to update the XML schema
- user friendly interface
- version control, validation

SVN

- SVN repository with validating commit hooks
- only valid XML documents are accepted by the server
- both UTF and Schema requirements must be met



The header of the temporary XML



The header of the TEI

Encoding

- temporary
- TEI XML with Manuscript-Linguistics Institute
- too complex for visualization (especially the transcription of letters)
- transform the TEI XML to TEI with XSLT

Two editors for modifying the transcription

Creators:

Stefan Jahn, David Dörner, David Pöhl, Jens Böttger



Manuscript-Linguistics Institute

Trainings

Throughout the year, ELTE.DH is offering free training opportunities for the faculty and the people in doctoral training programmes at ELTE. The main topics are:

Handling born digital materials

This training places emphasis on digital forensics and explains its methods which enable the restoration of the content of a hard disk, i.e. a born digital manuscript (including corrupted or deleted files). There are softwares that can reconstruct the writing process (corrections, transcriptions, erasures, etc.) to provide the genetic display of a particular work.

Publishing scholarly editions online

The publication of scholarly editions online is one of the most well-defined areas of digital humanities. Not only the methodology, the procedures and the markup language coding have been standardized over the last decades, but its reflexive academic practice has also emerged, with journals, conference series, manuals, and various trainings. This training provides insight into digital philology in practice-oriented ways, in the context of specific texts and services.



Digital science metrics and source text research

Participants become familiar with recently developed methods of measuring scientific performance in addition to traditional bibliometric tools. They get acquainted with other possibilities of mapping scientific networks and research collaborations beyond citation indexes, and with the expanding apparatus of impact assessment, including social media coverage analysis. All illustrated by a series of both national and international examples.

Digital cultural heritage

A wider audience, and even researchers increasingly interact with the objects of culture in digital form. However, the digital cultural heritage, while remaining connected with its analogue “origins”, functions primarily in accordance with the conditions of its own medium. These conditions are discussed in the training: the process of handling, archiving, publishing and describing cultural objects, whether books, manuscripts, works of art, or any other cultural artefacts which can be digitised; what happens with these objects in the collections, how they are gathered in portals such as Europeana, and what kind of possibilities and dangers the process of aggregation implies.

Stylometry: text analysis with computer methods.

Stylometry is an umbrella concept for the quantification of textual properties and conclusions drawn from these. These properties can be almost anything that occur in the text and their cardinality can be applied as a reference point: grammatical classes, word forms, punctuation, names, or even letters. The study of stylometric analysis serves to identify various patterns in a particular text: e.g. whether the text is verbal or nominal, what kind of lexical diversity it has, etc. Another approach to stylometry is drawing conclusions on the relationship between multiple texts, when, for instance, comparing texts in grammar or style. During the training, ELTE.DH provides practical insights into these two branches of stylometry.

Curriculum Development

ELTE.DH plans to increase the volume of common educational tools. Within the framework of the #dariahTeach project ELTE.DH is involved in development and translation of eLearning materials. ELTE.DH publishes the materials in Hungarian and uses it in the digital humanities education of ELTE.

DARIAH Teach was created to establish a platform and test modules for open-source, high-quality, multilingual teaching materials for the digital arts and humanities. A secondary, but not less important goal, is to strengthen alliances and foster innovative teaching and learning practices among members of the ESFRroadmap Digital Research Infrastructure for the Arts and Humanities (DARIAH).

The ERASMUS+ Strategic Partnership through which DARIAH Teach was created is under the lead of Susan Schreibman from Maynooth University (Ireland) and is collaborating with Aarhus University's DIGHUMLAB (Denmark), "Athena" Research Center (Greece), the Austrian Academy of Sciences (Austria), Belgrade Centre for Digital Humanities (Serbia), Erasmus University Rotterdam (The Netherlands), University of Luxembourg (Luxembourg), University of Lausanne (Switzerland) and the Swiss Institute of Bioinformatics (Switzerland).

The first advertising course material published in Hungarian translation was the Text encoding and the Text Encoding Initiative course, which was translated by ELTE.DH.



Guest Lectures

In the spring semester of 2017, ELTE.DH started preparations for the Centre with a series of lectures.

This meant a joint overview of the DH-related research in the various institutes, which helped the establishment of ELTE.DH, and was a professionally engaging and meaningful program for all of us. Moreover, there was a keen interest in the presentations of several researchers of the leading European digital humanities network, DARIAH.



Thorsten Ries, Ghent University (March 31, 2017, ELTE BTK,

Introduction to Digital Forensics: Hard Drive Philology) How does the philological comparison of manuscripts, correspondence, and text variations change in the digital age? In his presentation, Thorsten Ries discussed that the born-digital record, consisting of digital objects, temporary files, metadata, and fragmented traces of the writing process scattered across multiple system locations, has to be analyzed with regard to the specific historical computing context, its distributed materiality ensemble of hardware, operating system, and application. The exemplary analysis of several digital drafts and text fragments found on the hard drives of German poet Thomas Kling (1957–2005) sheds light on digital materiality from a practical digital forensic and critique génétique perspective.

Heinz Werner Kramski, Deutsches Literaturarchiv Marbach (March 31, 2017, ELTE BTK,

Workflow and some tools for preparing and ingesting unique digital objects (»born digitals«) on removable media) In his lecture “Long-Term Preservation of Born Digital Archival Objects: Serving the Unknown Future User”, Heinz Werner Kramski outlined from the Deutsches Literaturarchiv Marbach’s collection program a special responsibility to preserve unique digital holdings. These being »born-digital«, cannot be further reproduced and also cannot be redundantly acquired or cooperatively indexed. He explained, how DLA meets these growing quantitative and qualitative challenges and how it aspires to ensure the permanent preservation of authentic information objects and facilitate their use as a reliable long-term digital archive. He acknowledged that both the long-term

conservation of authentic usability, and appropriate, secure access remain great challenges for the digital preservation of holdings.

Henning Scholz, Europeana (April 5, 217, Petöfi Literary Museum)

Scholz presented, how Europeana Professional contributes to an open, knowledgeable, and creative society. Europeana offers access to academic research material and social media presence to professionals and institutions responsible for preserving digital cultural heritage. Their aim is to not only connect collections, but to create a functioning, future-oriented system of values as they recognize the huge value of linking researchers with cultural heritage institutions. With their programs, they bring Europe's galleries, museums, libraries, and archives within digital reach, and they monitor educational needs to develop useful features for education. The collection-based research is supported by a scholarship program.

ELTE lectures In the spring semester 2018

A new series of lectures demonstrated the vibrancy and breadth of work being done in the field at ELTE whilst simultaneously dealing critically with some of its issues and challenges.

2018-02-07 Media and communication. Teachers of the Department of Media and Communication, media researcher Katalin Orbán and journalist Attila Bátorfy talked about their research and the competencies improved by their education: the examination of existing media contents and databases is in fact an important cornerstone of their teaching work. Therefore, there are many ways and possibilities for collaboration between the department and ELTE.DH, in the use of plagiarism search systems, stylometry, faculty research repository and data visualization trainings.

2018-03-14 Slavic and Baltic Philology. Urkom Aleksander gave a lecture on the experiences and demands of the Slavic and Baltic Philological Institute and presented their website about the Slavic civilization. As an institution incorporating many foreign language departments, each with a small number of students, they actively use the on-line sources of these languages, although education is severely hindered by the lack of a proper hardware environment.

2018-04-12 Institute of Archeology. Dávid Bartus, associate professor of the Department of Ancient Archeology, presented the digital projects of the Institute of Archeology, the website of the institute, their blog, library database, online journal, online geoinformatics and photo database, as well as a database collecting archaeological sites. Parts of the pioneering but self-developed initiatives will certainly require conversion for future compatibility, thus the supportive presence of ELTE.DH will be important to them.

2018-04-26 Institute of Art History. The members of the Institute of Art History, Annamária Szőke, Mária Árvai, Júlia Katona and Ágnes Kusler introduced the collections and databases of the institute and presented which special softwares and integrated museum information systems serve educational purposes. Beside machine shortages, software support is the most urgent problem in their institute. Their database management software is outdated, the files needs immediate conversion, migration. Although an installed software would be necessary in the education of integrated systems, students can only practice by institutional visits, occasional licensees, and the trial version of the OMEKA system.

2018-05-08 Copyright implications. The deputy director of the Copyright Department in the Intellectual Property Office, Dr. Dénes Legeza described the issues of illegal publishing and the legal status of theses. He highlighted the new, now-evolving alternatives in digital publishing: the prospects for publishing commercially non-available books and unrestricted publication in an internal network.





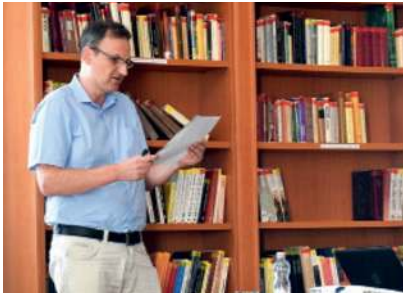
DH Budapest 2018

Digital humanities — as a new and quickly institutionalizing interdiscipline — serves as a new medium that connects various more or less independent disciplines we call(ed) humanities. This new “unity” in the digital medium — a medium, that was not designed to be used by scholars — takes place in a variety of ways: via digital cultural heritage aggregation, via using international data standards, mark-up languages, technologies such as linked open data, or through applying widely used CMS and repository software. The dual drive of the unstoppable curiosity to new digital methodologies and the urge to reflect on their influence theoretically has motivated us in launching a conference series. The Centre for Digital Humanities at the Eötvös Loránd University — in collaboration with DARIAH, CLARIN and Michael Culture Association — held the first of this series on 28–31 May 2018. DH_Budapest_2018 sought to provide a stimulating international forum to bring together researchers from Central Europe and beyond. It surveyed the current state of research in digital humanities in the hope of exposing further aspects of the role played by the digital medium in the present and the future of scholarly practices. ELTE.DH

invited submission of abstracts on subjects relating to digital humanities and social sciences ranging from practical approaches to novel methodologies and theoretical reflections concerning but not limited to the topics below:

- Digital scholarly editing
- Digital cultural heritage with a special focus on born digital documents
- Natural Language Processing (NLP), named entity recognition, topic modelling, big data analysis, text mining, semantic web technology etc.
- Past and present practices and networks of digital humanities in Central Europe
- Teaching digital humanities for students, teachers and researchers; eLearning

Five highly respected researchers from Europe and the United States accepted our request to hold a plenary lecture. More than 100 researchers took part in the conference from South Korea, the USA, Europe, as from Ireland and Hungary.





Workshops, presentations

DARIAH CEH workshop Budapest, ELTE.DH

Following the best practice of DARIAH Nordic Hub and inspired by an idea from a conversation with Mike Mertens (DARIAH CEO) in Budapest, April 2017, ELTE.DH organized the first meeting of Central European Hub with the support of DARIAH Humanities at Scale, 10th November 2017, Budapest.

This workshop has set up a work-plan of specific steps towards integration of DH efforts at the V4 level.

In the public session each country delegated a speaker who summed up the current situation of the digital humanities regional collaboration (with a special focus on DARIAH) and provided a short history of local DH research, infrastructure and other DH related activities (Austria, Czech Republic, Poland, Slovakia and Hungary).

This was followed by a closed session for founding DARIAH Central European Hub.

What are exactly the tasks of the DARIAH CEH? To raise awareness of DH knowledge and infrastructure, develop digital competences in the humanities, align DH teaching activities in Central Europe (#dariahteach, Course Registry), implement international Natural Language Processing tools for Social Sciences and Humanities, and pursue networking and fundraising.

ELTE.DH at Networkshop 2018 Conference, April 4-6, 2018, Eger

Gábor Palkó and Ádám Smrcz presented the outlines of the Online Knowledge Base of Hungarian Philosophy, an online library which consists of a repository storing contents such as texts, audio- and video files highly relevant from the perspective of the history of Hungarian philosophy. Some of its items are already available, but the Knowledge Base is continuously expanding. The other main component is a database enabling researchers to navigate between the contents via a strictly structured metadata system. The aim of the whole project is twofold: (1) to outline a possible canon of the history of Hungarian philosophy, while (2) also making it possible for researchers to determine the proper place of the contents included in the western history of philosophy. As for their presentation, they claimed that the latter effort will direct much more scholarly attention to the mostly neglected field of Hungarian philosophy in general, while also paving the way towards new paradigms in the historiography of Hungarian philosophy.



DARIAH-EU initiative at NTNU

Digital Research Infrastructure for the Arts and Humanities (DARIAH) held an open event at NTNU (Norwegian University of Science and Technology), Trondheim on 24 October 2017.

Goals: „Exploration of new ways of interactions between society and Information Communication Technologies (ICT) with a focus on the Humanities has the potential to become a key success factor for the values and

competitiveness of the Nordic region having in mind recent EU and regional political discussions in the field of Digital Humanities.” • Digital Humanities in the Nordic Countries: Current efforts, perspectives and challenges (Koraljka Golub et al., 2017).

Representing the ELTE DH Centre, Gábor Palkó introduced the projects of the ELTE.DH Centre and the plans of the DARIAH Central European Hub for regional cooperation, invited the representatives of the DARIAH Nordic Hub to the DARIAH event that was being held on 10 November 2017 in Budapest, where the representatives of the digital humanities institutes of the Visegrad 4 and Austria gave open presentations.

DARIAH-CZ Workshop on Digital Humanities 2018

Partners of DARIAH-CZ from the Czech Academy of Sciences, the Charles University and the National Library with support of the DESIR and LINDAT/CLARIN projects organised a one-day workshop to raise awareness about national and international activities in the field of digital humanities and to share experiences from European projects. The workshop took place at the Czech Academy of Sciences in April, 2018. At this workshop, Gábor Palkó gave a presentation about the history, current situation, and possible outcomes of digital humanities in Hungary, as well as about the activities and initiatives of ELTE.DH.

Digital Environments for Philosophy as a “Rigorous Science” Workshop at KU Leuven, 14 May 2018

The Husserl Archives organized a workshop on Digital Editions and Digital Research Environments for Philosophy on the occasion of the launch of the pilot project of Digital Husserl, the digital research environment dedicated entirely to the work of Edmund Husserl. The participants of the workshop sought to answer what digital scientific tools are needed today, particularly in philosophy, to keep up with current scientific procedures, as Edmund Husserl aimed to develop philosophy as a “rigorous science”. Two members of ELTE.DH, Gábor Palkó and Ádám Smrcz gave a presentation about the innovative concept of the Online Knowledge Base of Hungarian Philosophy project.

Real libraries - Reality of libraries III.

The motto of the Day of Hungarian Science 2017 was “Human centered Science”, which phrase emphasized the human factor as the central element of science, and the purpose of scientific knowledge, the service of mankind. The discussions highlighted the indispensability of library and information services in the rapid development of science, and the cooperation of science and society supported by library tools in order to influence the quality of life in a favorable way. The aim of the conference was to provide an opportunity both for national and cross-border representatives of the professional community to discuss the current issues of library and information science, to analyze the real and virtual role and possibilities of libraries, and to support libraries’ information mediating and research work. In the session of ELTE.DH, eight presentations demonstrated the value of bringing together the separate initiatives of digital humanities. Moreover, the range of research discussed in this session highlighted the growing shift away from digital fragmentation toward a digital unity within a university community.





Publications

International Journal of Digital Humanities (IJDH)

<http://www.springer.com/journal/42803/>

Springer, one of the most renowned publisher of scientific books and journals has started a new journal, the International Journal of Digital Humanities. The editor-in chief of the journal is Gábor Palkó, and the editorial team includes leading researchers from digital humanities centres in Vienna, Graz, Ghent and in Göttingen.

The first issue of IJDH is dedicated to the subject of “born-digital archives”. It ventures to shape the discussion on methodological approaches towards the digital archives of our historical present with philological, digital forensic, and archival science perspectives on private digital archives, domain-specific, as well as general web-archives, research archive infrastructures and hardware museums that preserve networking history. (Guest editor: Thorsten Ries)

The second issue of IJDH addresses the theory and practice of digital scholarly editing. The relatively new and mutable digital medium brings along its own set of challenges and opportunities. This field in flux is what the current issue of IJDH aims to come to terms with. Theoretical concerns are complemented with more practical approaches to digital scholarly editing, as researchers report on the development of new tools for modelling, processing, analyzing, visualizing variance and translations across versions. Through such juxtapositions, this issue of IJDH intends to offer an interesting cross-section of the current state of the art and inspire more debate in the field of digital scholarly editing. (Guest editor: Dirk van Hulle)

L. Varga Péter - Molnár Gábor Tamás - Palkó Gábor (ed.): *Digitális (szöveg)kultúrák a bölcsészképzésben*, ELTE BTK, Budapest, 2018. [Digital (Text) Cultures in the Humanities: a collection of essays]

In cooperation with the MTA - ELTE Digital Literacy and Literature Education research group, ELTE.DH published a volume on the connections of education and research in the humanities and digital media. 'Digital (Text) Cultures in the Humanities' includes the selected material of the workshop 'Digital Culture and Text Culture in Higher Education' held at the Faculty of Humanities at ELTE in January 2017, as well as new publications closely linked to the topic of the workshop. All studies argue the role of digital culture, its opportunities and challenges in recent education of humanities, both in terms of the discipline/ theory and the teacher training programme. Researchers and lecturers focusing primarily on literary, cultural and media studies, linguistics, and museology are seeking answers to the questions in the field of higher education, how the education of humanities and teacher training practices are transformed in the digital environment.

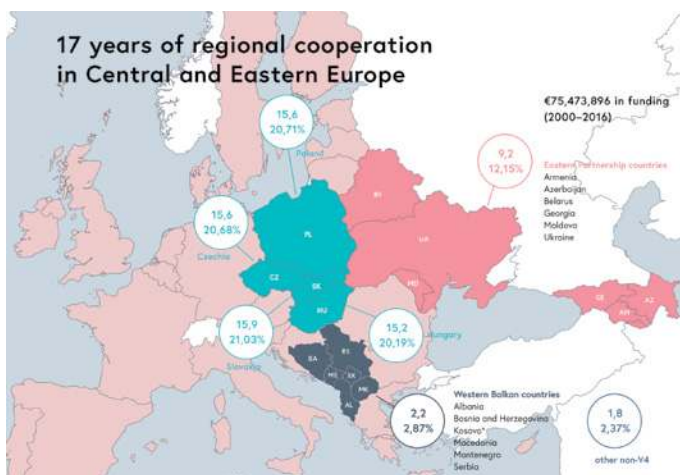


Grants

Distant Reading for European Literary History

ELTE.DH has joined the international COST Action research project “Distant Reading for European Literary History”. The aim of this COST Action is to create a diverse research network, jointly developing the resources and methods necessary to change European literary history. Based on the paradigm of distant reading (using computer methods for analysing large collections of literary texts) the COST Action creates a common theoretical and practical framework for innovative, sophisticated, data-driven computational methods for literary text analysis. COST Action promotes a wider, more coherent, and better foundation for European literary history and cultural identity by enabling cross-border, large-scale patterns and changes across European literary traditions to be perceived. It is paramount to create innovative ways of assessing, to analyze and compare the rich cultural production manifested in tens of thousands of literary works in many languages that are increasingly available in digital form, if Europe is to understand its cultural coherence, variety, and differentiations.

In this venture, all European languages and literatures, including the less-researched and less resourced ones, have an essential role to play. Consequently, building and coordinating the creation of a multilingual European Literary Text Collection (ELTeC) serve as a cornerstone of this action. The ELTeC core will contain at least 10 linguistically annotated subcollections of 100 novels comparable in their internal structure in at least 10 different European languages, including the Hungarian. The extended ELTeC will take the total number of full-text novels to at least 2,500. (Novels have been chosen among major literary genres for availability and size. Chronological limits are due to constraints related to copyright and availability of quality full texts.) ELTE.DH is currently taking part in the building of the Core Collection with the collaboration of the Hungarian Electronic Library (MEK) to encode and analyse Hungarian novels.



Visegrad Grant

ELTE.DH submitted a Visegrad Grant application under the project title 'Training Digital Scholars: Knowledge Exchange between V4 and Austria' in collaboration with seven partners (School of Computer Science of the Faculty of Mathematics and Physics, Charles University, Library of the Czech Academy of Sciences; Institute of Literary Research of the Polish Academy of Sciences; Supercomputing & Networking Center, Institute of Bioorganic Chemistry of Polish Academy



of Sciences; Centre for Information Modelling – Austrian Centre for Digital Humanities, University of Graz; Institute of Ethnology, Slovak Academy of Sciences; Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences; The Austrian Centre for Digital Humanities of the Austrian Academy of Sciences).

ELTE.DH and its partners will foster beneficial digital humanities (DH) practices and education in the V4 countries by organising a series of public tutorial and networking events, ie. three face-to-face training schools in digital humanities for researchers, teachers and students:

1. Linked Data and the Semantic Web for Humanities research, in Graz, Austria
2. Gentle Introduction into Natural Language Processing and Corpus Linguistics, Prague, Czech Republic
3. An introduction to Digital Philology, Budapest, Hungary

Partners

Governmental Information-Technology Development Agency (KIFÜ)
 HunCLARIN
 Hungarian Electronic Library (National Széchényi Library)
 Institute for Computer Science and Control (Hungarian Academy of Sciences)
 Institute for Literary Studies (Hungarian Academy of Sciences)
 Monguz Kft.
 Research Institute for Linguistics (Hungarian Academy of Sciences)

Members

Péter Kiszl (director)
 Gábor Palkó (co-director)

János Fodor
 Péter Horváth
 Balázs Indig
 Eszter Kovács
 Tünde Molnár
 Ádám Smrcz
 Alexandra M. Szabó
 Emma Takács

Former Members

Katalin Bella
 Dávid Halász
 Melinda Vásári

Address:

Múzeum krt. 6-8. 1088 Budapest Hungary

E-mail: digihum@elte.hu

Website: elte-dh.hu

Facebook: www.facebook.com/elte.dh/



science technology studies
real-time system
knowledge representation
language documentation
regional development
practice theory
do-it-yourself preservation
estate consciousness
knowledge management
semi-underground status
information technology
digital literary studies
digital dialectology
digital transformation
natural language processing
personal library
documentation
data collection
automatic metadata enrichment
digital classics
graph databases
social network analysis
open science
digital reflexivity
corpus analysis
digital archive
collaborative editing
dutch literature
quantitative formalism
nachlassbewusstsein
textualization
geonames
visualization
inscriptions
text mining
knowledge society
digital cultural heritage
digital media environment
oppositional
e-books
folklore texts
co-creation
dialects
archive
citizen science
communities
ethnomusicology
biographical dictionary
linguistic typology
data curation
language resources
new media
linked data strategy
infrastructure
text analysis
digital musicology
co-design
close reading
biography
annotation
academia-society collaboration
open access
ownership
metasearch
media history
NLTK tree
digital literacy
freemasonry
digital age
ottoman turkish
epidemic
sociology
digital writing
aerial archaeology
named entity recognition
alignment of vocabularies
linguistic geography
combinatoric reading
we-culture
food history
deep learning
software
digital participation
networks
search engine optimization
quantitative analysis
data aggregation
digital editing
born digital
open innovation
library and information science
tool development
parallax
findability
dialect geography
dbpedia and wikidata
digital storytelling
historical corpus
transylvania
history of science
digital scholarly editing
digitised dialect atlas
genealogical record linkage
teaching
public debate research
europeana entity collection
intangible cultural heritage
opposite-sex relations
natural language generation
data models
achievement competencies
digitization
open data
sentiment analysis
aerial photography
historical GIS
creative commons
inventory
place name
vowel formants
data archive
food cultures
biographical research
interactive
uratic languages
cultural practices
information retrieval
philosophy
digital platforms
non-monotonic logics
teaching