

eite.dh

# DH\_BUDAPEST\_2018

Centre for Digital Humanities – Eötvös Loránd University



# Contents

## POSTERS

### Judit Béres:

Bibliotherapy Database Promoting Mental Health & Wellbeing

4

### Richard Fejes:

Typology and Topology of an Alternate Reality Game

5

### János N. Fodor:

Geolinguistic methods in research of historical personal names. The AHSN project.

6

### Sarah Lang:

Fabula docet – Who wants sour grapes? Graz Repository of ancient Fables (GraF)

7

### Maria Niku:

The Elias Lönnrot Letters Online

8

### Yulia Petukhova:

Children and adults as amateur and professional authors:  
a comparative quantitative analysis

9

### Yevhen Rachkov:

University corporate culture investigation using a database technology  
(based on materials of Ukrainian classical universities)

10

### Christof Schöch:

Distant Reading for European Literary History. A COST Action

11

### Gergely Sipos:

The first building blocks of the European Open Science Cloud  
from the EOSC-hub project

13

### Karolj Skala:

DARIAH Thematic Service for Arts and Humanities in EOSC Hub

15

### Keily Suarez Rioja :

Panoramic Literature: Distant Reading and Data Visualization Techniques  
on Anthony Trollope's Literary Works

16

### Richard Zmëlik:

Fictional Semantics of Colours on the Background of  
Quantitative-Corpus Analysis

17

### Kinga Jelencsik-Mátyus – Eszter Simon:

HunCLARIN – Introduction

18

### Gábor Palkó – Zsófia Fellegi – Zsolt Czinkos – Réka Metzger:

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

19

### Gábor Palkó – Zoltán Kanász-Nagy – Anikó Mohay – Tünde Molnár:

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

20

## WORKSHOPS

### Bettina Fabos:

Duplicating the Fortepan Photo Archive Neighboring Countries

22

### Liviu Pop:

Building Memory Boxes: An Open Source Approach

23

### Sarah Tytler:

It Came from the Future! Integrating Digital Tools  
and Social Media into the Classroom

24

# POSTERS

# Judit Béres

University of Pécs

e-mail: beresjudit76@gmail.com

## **Bibliotherapy Database Promoting Mental Health & Wellbeing**

### **Keywords:**

Biblio-, Poetry Therapy / Reflective Reading  
/ Creative Writing / Mental Health Wellbeing /  
Biblio-, Poetry Therapy Materials / Database /  
Database Design

The poster aims to introduce a university project started in 2015, within the 2-year-long postgraduate bibliotherapy professional training in biblio/poetry therapy at the University of Pécs (Hungary). The 'Pécs School' of the Hungarian bibliotherapy education initiated a joint effort for the development of person-centered services linked to the interactive use of reflective reading and creative writing supporting people's personal growth, mental health and wellbeing. The most important outcome of this project is a bibliotherapy database which is under construction at the moment.

The database aims to connect people to helpful and inspiring books, highlights useful materials and provide mental health information for users, their families and practitioners. Our database gains inspiration from an American bibliotherapy database titled 'Reading for Recovery' (R4R) built by Judit Hajnal Ward, her colleagues and students at the Rutgers University in New Brunswick, USA. Our database follows the R4R model, but tends to cover more topics and target groups (not just substance abuse) with which we work and have multiple experiences. We plan to collect best practices and offer multilingual bibliotherapeutic descriptions of suitable and proved materials.

The main content of our database is based on the volunteer work of our students majoring in biblio/poetry therapy. This database can add to the reader's advisory toolkit of librarians, complement traditional therapies, and extend mental health support outside of traditional therapeutic settings. Students attending the Bibliotherapy Research seminar have written bibliotherapeutic descriptions about the interactive use of imaginative literature, non-fiction materials and creative writing, focusing on the support of people's personal growth, mental health & wellbeing. At the moment we have more than 500 descriptions as a base for a future bibliotherapy database, of which technical building is in progress.

### Project aims:

- provide theme-appropriate, personalized, annotated information for individuals, parents, teachers, librarians, and helper professionals who apply individual and group bibliotherapy
- highlight quality literature of all kind as a life-support system, promote reflective reading and creative writing as mental health tools that are available for anybody
- build international links through items relevant/available in national contexts and languages

### Database design plans:

1. Detailed annotations: recommended readings based on real trials & written by trained bibliotherapists
2. Structure of the descriptions: Book byte/ Suggested readers/ Therapeutic insights/ Bibliographic data (for printed and online materials)
3. Search options by: author/ title/ genre/ topic/ target group
4. Further useful information: URL for materials that are available online/ trial experiences/ methodological toolkit/ help desk

## Richard Fejes

Eötvös Loránd University,  
Budapest, Hungary

e-mail: afejesrichard@gmail.com

### **Typology and Topology of an Alternate Reality Game**

#### **Keywords:**

ARG / Alternate Reality / Digital Storytelling /  
Pervasive Gaming / Hybrid Literature

This ongoing research is focusing on ARGs (or Alternate Reality Games) which are an emergent form of hybrid narratives. These pieces are simultaneously operating on different platforms at the same time, interacting with social media platforms as well as offline surfaces such as books or even billboards. The mechanics of an ARG is based on its medial reach so a comprehensive story can only be formed if the fragmented storyline is read or played on multiple platforms, including our everyday reality itself. This is where the alternate reality part comes in, as these works of art are penetrating and permeating our level of perception, our empirical world.

This current presentation on this ongoing project would like to map out a single piece of ARG, the mainly YouTube based EverymanHYBRID series. With the help of this narrative mapping could help us develop a typology of the semantic, medial and transtextual elements of this kind of storytelling. The expectations are that what we would find describing and analyzing the hybrid nature of the Alternate Reality Games would help us further to reflect on the nature and techniques of contemporary (progressive) literature.

## János N. Fodor

Eötvös Loránd University,  
Budapest, Hungary

e-mail: [nfodor.janos@btk.elte.hu](mailto:nfodor.janos@btk.elte.hu)

### **Geolinguistic methods in research of historical personal names. The AHSH project.**

#### **Keywords:**

Geolinguistics / Historical Surnames /  
Namegeography

The main goal of the poster is to introduce the project of Atlas of Historical Surnames in Hungary (AHSH).

Linguistic geography has provided linguistic research with a general tool whose origins date back to the 19th century, within both Hungarian and international standards. According to its broad definition, geolinguistics refers to “spatial linguistics,” a discipline examining those influences brought about by contact between languages, the regional variations existing within a language and the special properties possessed by dialects. As a method, linguistic geography enables researchers to depict the spatial movements produced by contact influences between languages and dialects, thereby allowing for the effective recreation of diachronic linguistic processes and their connections.

For representing the data on map, it was by all means required to develop computer software which is able to perform different linguistic tasks (e.g. generating morphological maps) considering the database’s specifics. The database management software of geographic information system generates cartograms pursuant to calculations of two sorts (absolute and relative distribution) that show the regional features of names, name types and language elements (e. g. suffixes).

As a result of the research in geonomastics—begun in 2010—numerous advances have occurred not only in analyzing various aspects of the name data on which this project is based, but also in building a digitalized database and in creating and developing the own software necessary for generating cartograms.

## Sarah Lang

University of Graz,  
Centre for Information Modelling,  
Graz, Austria

e-mail: sarah.lang@uni-graz.at

### **Fabula docet - Who wants sour grapes? Graz Repository of ancient Fables (GraF)**

#### **Keywords:**

GAMS Repository / Graz Repository of Ancient Fables (GraF) / Digital Schoolbook

Fables are short and easy to translate and thus constitute a great medium for introducing pupils to literary studies in Latin and Ancient Greek class. They have, however, been frowned upon by science as “children’s stories”, so authors like Phaedrus, Avianus and Babrius haven’t been treated much so far.

This Sparkling Science project aims at combining the educational aspects of fables that make them valuable for use in schools as well as the scientific aspect of fable scholarship. In doing so, it introduces pupils to scientific methods in the creation of a digital (school book) edition of selected fables via a medium they are already familiar with, as a joint effort of students and scientists alike.

Providing the original text in TEI annotated form - scientific comments as well as simple vocabulary help - the repository (available via the FEDORA based asset management GAMS) is completed with didactic materials as well as scientific introductions. It thereby challenges the Digital Humanities to explore this new kind of medium - somewhere between a digital (scholarly) edition and a learning platform.

The digital edition (“fable repository”) is the centrepiece of the project and contains a selection of fables (Phaedrus, Avian, Babrius, Aesop) including an apparatus, providing enrichment of the original texts with vocabulary, translations, explanations and parallel texts. Thereby fables are treated in a scientifically correct way - as literary works and not just children’s stories - but the project also makes use of the unique accessibility and rich imagery they offer for teaching ancient languages. The goal of the project is not to provide an interactive learning platform but to innovatively explore the concept of the “digital school book”, created by the pupils themselves but grounded in up-to-date scientific research on fables, provided by the scientists coordinating the project.

## Maria Niku

Finnish Literature Society,  
Helsinki, Finland

e-mail: [maria.niku@finlit.fi](mailto:maria.niku@finlit.fi)

### The Elias Lönnrot Letters Online

#### Keywords:

Elias Lönnrot / Digital Editions / Digital  
Publishing

The correspondence of Elias Lönnrot (1802–1884, doctor, philologist, creator of the national epic *Kalevala*) comprises of 2 500 letters or drafts written by Lönnrot and 3 500 letters received. Elias Lönnrot Letters Online (<http://lonnrot.finlit.fi/omeka/>), first published in April 2017, is designed for scholars and general public interested, apart from Lönnrot himself, in the Finnish 19th century nationalistic academic community and in the study of epistolary culture. The correspondence offers source material for research in biography, folklore studies and literary studies; general history, medical history and the history of ideas; the study of ego documents and networks; and corpus linguistics and history of language.

As of January 2018, the edition contains about 2000 private letters and drafts written by Lönnrot. The official letters, such as the medical reports submitted by Lönnrot in his office as a physician, will be added during 2018. The final stage will involve finding a way of publishing for the 3500 letters that Lönnrot received.

The edition is built on the open-source publishing platform Omeka. Each letter is published as facsimile images and an XML/TEI5 file, with metadata and transcription. The letters are organised into collections according to recipient. A faceted open search powered by Apache's Solr, which covers the metadata and transcriptions, allows limiting the initial search by collection, date, language, type of document and writing location.

The guiding principle of Elias Lönnrot Letters is openness of data. All the data is made openly available.

Researchers and other users are free to download and modify the XML/TEI5 files. The users can download the XML/TEI5 files of all the letters, or of a smaller section. The feature is also integrated in the faceted search. Thus, a researcher can download the XML files and study them for example with the linguistic tools provided by the Language Bank of Finland. Similarly, the raw data is available for processing and modifying by researchers develop digital humanities tools and methods to solve research questions.

The transcriptions can be downloaded as plain text. This format is needed for qualitative analysis tools like Atlas. In addition, researchers in humanities who do not need XML files will benefit from the ability to store relevant data in an easily readable format.

Users can export the statistical data contained in the facet listing of each search result for processing. Statistical data like this is significant in handling large masses of data, as it can reveal aspects that would remain hidden when examining individual documents. A researcher may want to know when and with whom Lönnrot primarily discussed a given theme. The statistical data readily reveals such information, while compiling such statistics by manually going through thousands of letters would be an impossibly long process.

The easy availability of data in the online edition will hopefully foster collaboration. The SKS is already collaborating with the Language Bank, which has received the XML/TEI5 files. A section of the letters was given to the international READ project, which w to develop machine recognition of old handwritten texts.

## Yulia Petukhova

Ruprecht Karl University of Heidelberg,  
Heidelberg, Germany

e-mail: avvrik@gmail.com

### **Children and adults as amateur and professional authors: a comparative quantitative analysis**

#### **Keywords:**

Quantitative Analysis / Amateur Writing / Children as Authors / Topic Modelling / Figurative Language / R Statistics

While some books written by amateurs and teenagers are published as hard copies, other literary works remain on the web forever. Do the latter suffer from the lack of managerial effort of their authors or are there formal features which make them unfit for the book trade? It is assumed that learning creative writing helps eliminate amateur problems and acquire the techniques of professional writers. In this project, I would like to challenge the creative writing theory and see if those tools are what in fact distinguishes an amateur from a professional.

My doctoral research is aimed at finding formal characteristics of amateur adults' and children's writing which differentiate them from that of professional authors. This poster will show the results of the comparative analysis of one of the literary genres, poetry, authored by three groups of people: children and teenagers (age 5 to 18), adult amateur poets, and professional poets who are generally thought of as targets of contemporary poetry.

The method of the analysis involves the use of Python in order to parse the websites with the poems in question and create three corpuses for lemmatization and POS-tagging, which prepares the texts for statistical analysis by the relevant features, derived mainly from creative writing textbooks. These features include, for example, frequency of adverb usage or frequency of the use of 'adjective plus noun' pattern, which are regarded as simplistic when it comes to poetry in creative writing classes. The question is whether professional authors are, in fact, characterized by a lower percentage of these parameters, which was also accessed using the R statistics.

An essential method used in this project is topic modelling. Defining the semantic structures hidden in the texts provides the opportunity to see what concerns different groups of people divided by age and level of proficiency. It allows to see the trends that exist among younger writers who are much more uniform in the choice of topics than professional poets. While adult authors use indirect ways of approaching a topic, teenagers are more straightforward in conveying a message.

One of the major challenges of this project is the quantitative analysis of figurative use of language. Currently, there are no efficient computational methods of approaching the majority of figures of speech. The use of figurative language is an important feature that differs in children's and adults' writing. While its statistical analysis would be invaluable for this project, computer recognition of metaphors is an area that requires further research. In this project, it is partially tackled by examination of semantic trees. Children tend to stay within the same topic throughout the whole poem while adult professional writers incorporate unexpected semantic groups which appear to be juxtaposed with the branches of the main topics.

This research is a work in progress which is supposed to elucidate the underlying features that make a professional writer and those features that can strike the eye as childish or amateur.

## Yevhen Rachkov

V.N. Karazin Kharkiv National University,  
Kharkiv, Ukraine

e-mail: [yevhen.rachkov@karazin.ua](mailto:yevhen.rachkov@karazin.ua)

### **University corporate culture investigation using a database technology (based on materials of Ukrainian classical universities)**

#### **Keywords:**

Ukrainian Classical Universities / Symbol /  
Emblem / Database / Content-Analysis

During several centuries of Ukrainian classical universities existence, their symbols and emblems have composed a special complex of visual sources. It can demonstrate the specific features of a university corporation and indicate the existence of a university identity in the modern period. Symbols and emblems are not only an attribute of ritualized life of universities. First and foremost, they are the signs of university tradition as well as a language which is used in order to read a “cultural code” of universities.

The objectives of this work are to retrace a process of formation and evolution of symbols and emblems of Ukrainian classical universities during all time of their existence and to find out the specifics of self-reflection and self-representation processes of university community via an analysis of their symbols and emblems.

In order to attain this objective, we perform the content-analysis of symbols and emblems of 9 classical Ukrainian universities located at Chernivtsi, Dnipropetrovsk, Donetsk, Kharkiv, Kyiv, Lviv, Odesa, Simferopol, and Uzhhorod. We construct a relational database in the database management system Microsoft Access 2010. The database nucleus (basis) consists of 6 tables: “Emblems”, “Figures”, “Sources”, “Universities”, “Text”, and “Use”. The database includes also auxiliary tables which ensure an operation of main tables. We analyze the external and internal (meaningful) descriptions of university symbols and emblems. First ones are the fields with information about attributive signs of images, their form, formal status, specificity of use etc. Second ones are the fields with information about figures (elements of images), type of symbols (international, state, regional, corporate), color gamut etc. The database consists of about 400 images of symbols and emblems of Ukrainian classical universities. Major part of materials has been collected by the author in museums of history of these universities.

Because of unsystematic and sporadic character of university symbols and emblems origin, they are mostly inaccessible for analysis. Therefore it is quite complex to evaluate a percentage of collected materials in parent population.

The presented database is the first attempt to collect, to systematize, and to analyze the symbols and emblems of Ukrainian classical universities during all time of their existence. We have created the database which allows to observe the transfer of traditions of West European universities to Ukraine and to understand the specific character of self-reflection and self-representation processes of universities located in different regions of Ukraine.

# Christof Schöch

University of Trier,  
Trier, Germany

e-mail: [schoech@uni-trier.de](mailto:schoech@uni-trier.de)

## **Distant Reading for European Literary History. A COST Action**

### **Keywords:**

Corpus building / Text Analysis / Literary  
History / Literary Theory / Computational  
Linguistics / Community / Network

### **Authors**

Schöch, Christof; Eder, Maciej; Odebrecht, Carolin; Kestemont, Mike; Primorac, Antonija; Tonra, Justin; Mihurko Poniž, Katja; Kanellopoulou, Catherine.

### **Introduction**

This poster aims to stimulate awareness of the existence of the newly-established COST Action on “Distant Reading for European Literary History” (2017-2021). In the context of this networking project, “distant reading” is understood as an umbrella term for recent computational, and particularly quantitative, approaches to the study of large collections of texts. This paradigm is here applied to the multilingual literary traditions of Europe in the long nineteenth century.

### **What is a COST Action?**

COST ([www.cost.eu](http://www.cost.eu)) stands for ‘European Cooperation in Science and Technology’: COST Actions are essentially networking initiatives focused on a particular, timely and innovative research topic, aiming to bring together a critical mass of researchers from Europe and beyond. COST Actions coordinate their activities through working group meetings and offer Training Schools and opportunities for scientific exchange. Examples of previous COST Actions in Digital Humanities include Interedition (<http://www.interedition.eu/>, 2008-2012) and e-Lexicography (<http://www.elexicography.eu/>, 2013-2017).

### **Aims of the “Distant Reading” Action**

We aim to create a vibrant and diverse network of researchers jointly developing the resources and methods necessary to change the way European literary history is written. Fostering insight into cross-national, large-scale patterns and evolutions across European literary traditions, we will facilitate the creation of a broader, more inclusive and better-grounded account of European literary history and cultural identity. We will foster distributed research, the systematic exchange of expertise, and the visibility of all participants, activities and resources.

In terms of scientific objectives, we will coordinate the creation of a multilingual European Literary Text Collection (ELTeC). It brings together comparable sets of nineteenth-century novels from at least 10 different European languages. The purpose of the ELTeC is to serve as a benchmark corpus for the evaluation and development of annotation tools and distant reading methods across languages and as the basis for investigations into patterns and trends in literary history in multiple literary traditions. We will also use the ELTeC to establish best practices and develop innovative methods of Distant Reading for the multiple European literary traditions.

Furthermore, we will engage in an investigation into the theoretical consequences of Distant Reading approaches for literary history and literary theory. We also aim to foster the acquisition of state-of-the-art methods related to data curation, standards, best practices and quantitative analysis in workshops and training schools. Last but not least, we aim to address the current gender imbalance among practitioners of Distant Reading research.

#### **The network**

Our network of members is currently comprised of researchers in Corpus Linguistics, Computational Linguistics, (Digital) Literary History and Literary Theory from 27 different countries and more than 40 cities across Europe and beyond.

#### **Learn more, learn how to join**

To learn more, see the Action's profile page [http://www.cost.eu/COST\\_Actions/ca/CA16204](http://www.cost.eu/COST_Actions/ca/CA16204) and the full proposal linked there ("Memorandum of Understanding"). The Action's website will soon be available at <http://www.distant-reading.net>. Researchers from Computational Linguistics, (Digital) Literary Studies as well as Computer Scientists and Librarians are welcome to get involved!

## Gergely Sipos

EGI Foundation,  
Amsterdam, Netherlands

e-mail: [gergely.sipos@egi.eu](mailto:gergely.sipos@egi.eu)

### **The first building blocks of the European Open Science Cloud from the EOSC-hub project**

**Keywords:**  
EOSC open science

EOSC-hub is a 33 million Euro H2020 project that started in January 2018 with the involvement of over 100 institutes. In the next three years the project will establish the first elements of the European Open Science Cloud (EOSC) – the infrastructure aimed to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market.

This poster introduces the EOSC-hub project and its relevance to digital humanities communities. EOSC-hub will define, create and operate the integration and management system of the EOSC. This integration and management system (the Hub) builds on mature processes, policies and tools from the leading European e-infrastructures to cover the whole life-cycle of services from planning to delivery.

Through this management system online and ‘human’ services, software and data will be delivered towards researchers via a single service Marketplace. The EOSC-hub consortium already includes providers from 3 e-infrastructure communities (EGI, EUDAT, INDIGO-DataCloud), and from 18 Research Infrastructures (mostly from the ESFRI Roadmap). The group of service providers is expected to radically grow in the next years.

The Hub acts as a single contact point for researchers and innovators to discover, access, use and reuse a broad spectrum of services for advanced, data-driven research. The project will also improve skill and knowledge among researchers and service operators by delivering specialized trainings and by establishing ‘competence centres’ where existing services can be tested, new services can be co-created with the involvement of scientific user communities. In the area of engagement with the private sector the project will run a ‘Digital Innovation Hub’ to stimulate an ecosystem of industry/SMEs, service providers and researchers in the form of ‘business pilots’ that aim to produce services for commercial markets.

CLARIN and DARIAH already participate in the project from the Digital Humanities discipline. Specific service provider institutes from these two ESFRI Research Infrastructures will deliver science discipline specific services through the Hub for the benefit of disciplinary and inter-disciplinary researchers and innovators worldwide.

CLARIN Thematic Service providers will setup and operate the component metadata framework (CMDI) as a way to stimulate the discoverability of data sets, making the citation of these more convenient and to take away the barriers to automated processing of data. This will be achieved by developing a uniform and robust workflow to (1) gather metadata descriptions from the various scientific communities linked to CLARIN, (2) convert these files into a suitable CMDI-based equivalent, (3) perform a highly-parallelised indexing of all the resulting metadata files. The CLARIN providers will conduct training and dissemination about CDMI.

DARIAH Thematic Service providers will operate the 'DARIAH Gateway', a web portal designed to cater for the digital data analysis needs of digital humanities researchers. The gateway already offers several cloud-based services and applications to DH researchers: Semantic and Parallel Semantic Search Engine, DBO@Cloud, Multi-Source Distributed Real-Time Search and Information Retrieval Application (SIR), Workflow Development and File transfer. The DARIAH providers will extend these and will support other closely related research communities coming from arts, humanities and social sciences by training and technical support.

## Karolj Skala

Ruđer Bošković Institute,  
Zagreb, Croatia

e-mail: skala@irb.hr

### **DARIAH Thematic Service for Arts and Humanities in EOSC Hub**

#### **Keywords:**

Cloud Service / E-Infrastructure / Integrated Gateway

The “grand vision” for the DARIAH Research Infrastructure is to facilitate long-term access to and use of all European humanities and cultural heritage information for the purposes of enhancing and expanding research, thereby increasing our knowledge and understanding of our history, heritage, languages and cultures.

DARIAH activity based on digital humanities, includes the systematic use of digital resources, storage, computing and software in the arts and humanities, as well as the reflection on their application.

The DARIAH Science Gateway serves as a central point for accessing different computing and storage e-infrastructures (e.g. EGI FedCloud), as a place where applications, tools, datasets and repositories are shared among different users and research projects. The current gateway is already integrated with the EGI infrastructures (grids and clouds). The services provided via the EGI DARIAH CC Gateway are cloud-based services that are running on the EGI FedCloud compute and storage resources. During the EGI DARIAH CC (EGI-Engage project) a DARIAH virtual organization has been established.

Currently, the gateway is integrated only with the EGI infrastructure and the EOSC-Hub project plan is to integrate the gateway with the EUDAT data resources as well as with selected INDIGO services in uniform Service catalog as a DARIAH Thematic Service.

The EOSC-Hub mission is to contribute to the EOSC implementation by enabling seamless and open access to a system of research data and services provided across nations and multiple disciplines. The project will offer the resources via the Hub – an integration and management system of the European Open Science Cloud, acting as a European-level entry point for all stakeholders. The Hub will deliver a catalogue of services, software and data collections from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-Infrastructures. The Hub builds on mature processes, policies and tools from the leading European federated e-Infrastructures to cover the whole life-cycle of services, from planning to delivery. The Hub aggregates services from local, regional and national e-Infrastructures in Europe and worldwide.

Service providers of the DARIAH Science Gateway are:

Ruđer Bošković Institute (RBI), Zagreb, Croatia

Hungarian Academy of Science, Institute for Computer Science and Control (MTA SZTAKI), Hungary;

Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG), Germany;

Communications: info@irb.hr

Relevant links:

[https://wiki.egi.eu/wiki/Competence\\_centre\\_DARIAH](https://wiki.egi.eu/wiki/Competence_centre_DARIAH)

<http://www.eosc-hub.eu/>

## Keily Suarez Rioja

Catholic University of Leuven,  
Leuven, Belgium

e-mail: keily.s.rioja@gmail.com

### **Panoramic Literature: Distant Reading and Data Visualization Techniques on Anthony Trollope's Literary Works**

#### **Keywords:**

Distant Reading / Digital Humanities / Literature  
/ Literary Analysis / Corpus Linguistics /  
Data Visualization / Trollope / Computational  
Analysis

The study of Literature calls for the study of literariness as well as the study of its textual data. More traditional methods of literary analysis, such as 'close reading,' fulfill this call to a certain extent. However, the sole use of traditional methods does not seem to be sufficient or efficient enough to answer all questions concerning Literature, literariness, and literary texts. In the last decades, data analysis of texts has proven to be advantageous and 'distant reading' techniques have posed as a solution to answer those questions that close reading could not answer. It is through research and discoveries in the field of Digital Humanities that the digital can now be incorporated in the study of Literature.

This project proposes the use of digital tools to create a qualitative and quantitative substantial illustration of what can result from the use of corpus linguistics methods and data visualization techniques when used to analyze phenomena in the literary realm. The project and DH Master Thesis topic (KU Leuven, supervised by prof. dr. Katrien Verbert, prof. dr. Ortwin de Graef, and prof. dr. Dirk Speelman), involves the analysis of Anthony Trollope's literary works through computational applications and data visualization of the results found and/or generated. The corpus used in the project mainly focuses on Trollope's literary works, however, works by contemporary authors, such as Charles Dickens and William Makepeace Thackeray, are used in order to determine Trollope's style of composition and narrative characteristics through computer-detected comparisons.

Some examples of results found through the running of corpus linguistics analysis applications in the R package 'stylo' --which measures the stylometrics in a corpus-- to the corpora mentioned above are Trollope's distinctive syntax ("at any rate"), vocabulary preference and avoidance ("declared," "cried"), and style or genre clustering; all presented in a visually comprehensive form.

The project aims to provide a more "panoramic" view of certain literary aspects that only distance can provide. More advanced methods of visualization, such as an interactive model, will complement the project. The visualizations of the data gathered from digital analyses aim to offer a clearer view of the whole as well as serve as a means for further research, as such techniques often lead to new discoveries and new questions.

The resulting data is to be paired with close reading techniques in order to produce in-depth results about Trollopean literature and understand the subject more exhaustively.

The overall end goal of the project is to show how computational techniques used in the field of Digital Humanities can in fact contribute to the field of Literature and Literary Studies.

## Richard Změlík

Palacký University Olomouc,  
Olomouc, Czech Republic

e-mail: richard.zmelik@upol.cz

### **Fictional Semantics of Colours on the Background of Quantitative-Corpus Analysis**

#### **Keywords:**

Electronic Corpus of Prosaic Works / Fictional Semantics of Colours / Lexical-Statistic Models / Topic-Statistic Models / Literary-Theoretical Interpretation of Models

Between 2012 and 2015, an electronic corpus of prosaic works of Jan Čep, a Czech writer (1902–1974), was assembled. Based on this corpus, research focused on measuring of so-called thematic concentration of text was realized, analysing Čep's three earliest collections of short stories. The research generally and primarily aimed at finding out and testing to which extent quantitative-corpus methods can be used in literary theory and in which way results of such methods can be meaningfully interpreted by literary theorists. Particular results were published in the monograph entitled XYZ (2015).

Current research uses the same material. It focuses on fictional semantics of colours and utilises the above mentioned corpus sources. Linguists explain the semantics of elementary names of colours also from the cognitive perspective, as results of conceptualisation processes (Wierzbicka) the fundamental frame of which is formed by the natural world, a perceiving subject, and historical-cultural context (Waszakowa). In linguistics, semantics of colours in literary texts is explained as so-called textual meanings. These meaning are derived from systemic meanings that are based on conceptual meanings (Tokarski). In analysis of textual meanings a wide scale of such meanings is observed as a correlative of artistic intention of texts.

We intend to specify general systemic qualities of the semantics of colours having its source in their conceptualization within the frame of a particular set of fiction worlds that represent Jan Čep's proses. For these purposes we use elementary lexical-statistic models together with quantitative-corpus analysis of contexts (collocations) that provide a comparative view of the stratification of concepts for elementary colours, appearing not only in the analysed works of Jan Čep.

The ways of meaningful incorporation of statistic models in literary-theoretical interpretation is an important issue that we study again within this research.

In the poster we will present statistic models that are based on quantitative-corpus treatment of Jan Čep's proses as well as on selected prosaic corpora of other Czech writers (Bohumil Hrabal, Karel Čapek, Karel Hynek Mácha). We will demonstrate the form of stratification of elementary colours in the works of individual authors in the context of asymptomatic stratification of elementary terms signifying primary colours (see Berlin – Kay) in Czech (Schmidtová). We namely intend to focus on the situation in the prosaic work of Jan Čep and to present its latent structural mechanisms that participate in conceptualization of fiction semantics of elementary notions for colours, on second degree conceptualization (as we call it).

# Kinga Jelencsik-Mátyus – Eszter Simon

Research Institute for Linguistics,  
Hungarian Academy of Sciences  
Department of Language Technology  
and Applied Linguistics  
Research Group for Language Technology  
Budapest, Hungary

e-mail:

matyus.kinga@nytud.mta.hu

simon.eszter@nytud.mta.hu

## HunCLARIN – Introduction

### Keywords:

Linguistics / Natural Language Processing /  
Infrastructure

HunCLARIN is a strategic research infrastructure-group of Hungary's leading knowledge centres involved in R&D in speech- and language processing. It contains linguistic resources (different types of corpora and NLP tools) that form the basis of research. The infrastructure-group has obtained an "SKI" qualification (Strategic Research Infrastructure) in 2010, and has been significantly expanded since. Currently comprising more than 40 resources, the project includes several general- and specific-purpose text corpora, different language processing tools and analysers, linguistic databases as well as ontologies.

Research Institute for Linguistics (Hungarian Academy of Sciences = RIL HAS) is the coordinator of the HunCLARIN project. Altogether, HunCLARIN has 8 member institutions, providing state-of-the-art resources for researchers working with larger texts.

RIL HAS was a co-founder of the European CLARIN project, which aims at supporting humanities and social sciences research with the help of language technology and by making digital linguistic resources more easily available. CLARIN's coordination body is an ERIC (European Research Infrastructure Consortium), an international legal entity, established by the European Commission in 2009. CLARIN ERIC's members are governments or intergovernmental organisations. CLARIN ERIC currently has 20 members and 2 observer countries.

In accordance with the goals of CLARIN ERIC, HunCLARIN is working on making the research infrastructures developed by the respective centres directly accessible for researchers through a common network entry point. A general goal of the infrastructure is to realise the interoperability of the collected research infrastructures and to enable comparing the performance of the respective alternatives and to coordinate different foci in R&D.

For more information please visit [clarin.hu](http://clarin.hu)

**Gábor  
Palkó –  
Zsófia  
Fellegi –  
Zsolt  
Czinkos –  
Réka  
Metzger**

Petőfi Literary Museum  
Budapest, Hungary

e-mail:

palko.gabor@pim.hu

fellegi.zsofia@pim.hu

czinkos@gmail.com

metzger.reka@gmail.com

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

**Keywords:**

Text Encoding / Literary Database /

Digital Scholarly Editing

The joint project of the Petőfi Literary Museum and the Hungarian Academy of Sciences Institute for Literary Studies named DigiPhil (Online knowledge base of scholarly text editions, bibliographies, and databases) in the framework of the Arany János Memorial Year has developed a virtual research environment. This environment provides 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 accelerates encoding and data enrichment while providing software and hardware independence on the client side. In the same time, it produces synthetically correct codes on the server side in a secure data environment through functions as versioning and automatic error correction (validation). The codes produced are then 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 Arany János Critical Edition.

**Gábor  
Palkó -  
Zoltán  
Kanász-Nagy -  
Anikó  
Mohay -  
Tünde  
Molnár**

Centre for Digital Humanities  
Budapest, Hungary

e-mail:

zkanasznagy@monguz.hu  
mohaya@pim.hu  
tndemolnar@gmail.com  
palko.gabor@btk.elte.hu

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

**Keywords:**

Repository / Semantic Web / Metadata

The Centre for Digital Humanities at Eötvös Loránd University is creating a new Islandora Claw based online repository in order to store and preserve content according to the most recent technological developments. The pilot project is the Hungarian Philosophical Database which 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.

The software base for this and other projects in the future is the open source Islandora CLAW which was released not that long ago and had been greatly improved compared to previous versions. The most essential upgrade was the change in the infrastructure. Islandora CLAW consists of a Drupal 8 content management system, a Fedora 4 database, a Blazegraph database, a Solr search engine and a synchronisation component.

Some of the features provided by Islandora CLAW 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.

The Islandora Claw is still very much in development, which means that some potentially useful components might be missing but are continuously developed and then added.

# WORKSHOPS

## Bettina Fabos

fortepan.hu, Fortepan,  
Budapest, Hungary  
fortepan.us, University of Northern Iowa,  
Cedar Falls, Iowa, United States

e-mail: [fabos@uni.edu](mailto:fabos@uni.edu)

### Duplicating the Fortepan Photo Archive Neighboring Countries

#### Keywords:

[Archive](#) / [Photographs](#) / [Family Snapshots](#) /  
[History](#) / [Timeline](#)

Fortepan ([fortepan.hu](http://fortepan.hu)), which originated in Hungary in 2010, is a public, Creative Commons (CC)-based archival framework that displays and shares historical 20th century photographs. The curated photographs are primarily from Hungary, most are from personal family collections, and all come from collections that would otherwise remain hidden from public view or lost altogether. The photos are arranged along a timeline that immediately situates them in a specific moment in Hungarian history. The project takes visitors through a rich and complicated visual landscape designed to generate dialogue about the country's social and cultural values, change and continuity, and the evolution of identities and environments in Hungary's past.

The archive's deep-seated public CC orientation, based on sharing and volunteerism, begins with the everyday people and institutional archivists who allow their images to be collected, curated, scanned, and shared freely with the larger public. The first Fortepan photos came from co-founders Miklós Tamási and Ákos Szepessy's personal collection, which they developed by rescuing unclaimed family snapshots from dumpsters and attics and mining their own personal family albums. Then came donations from their friends' families, followed by hundreds of interested individuals wanting their own personal photographs to be part of an exciting new bottom-up historical record. Since 2015, even GLAM institutions throughout Hungary have sought to benefit from Fortepan's popular success as an online platform and community forum, and have dedicated a portion (or in some cases, all) of their collections to the public CC archive. Fortepan currently shares over 100,000 high-resolution, instantly downloadable, easily searchable photographs with the CC BY license.

As a public CC collection that uniquely juxtaposes family snapshots with professional photographs, Fortepan invites the general public (as well as educators, researchers, and artists) to develop alternative histories counter to the hegemonic narratives that so often dominate historical discourse. Another key element of Fortepan's success has been the archive's ability to build community around its photographic mission of a shared public commons. Fortepan has established an avid network of donors, volunteers, online contributors, and artists who create inspiring projects based on Fortepan images. The Fortepan "concept" is also ripe for replication in other countries, a process that has already begun.

In 2015, a team at the University of Northern Iowa (UNI) led by Dr. Bettina Fabos launched Fortepan Iowa ([fortepan.us](http://fortepan.us)). The new sister site replicated Fortepan except with a focus on everyday family snapshots from Iowa. As our teams continue to collaborate, we hope to facilitate a global community of Fortepan archives. Our workshop will be sharing what we know, promoting further collaboration with other countries, and supplying interested parties with protocol toolkits and front/backend code.

## Liviu Pop

Romanian Academy,  
Institute of Folklore Archive,  
Bucharest, Romania

e-mail: [liviu.o.pop@gmail.com](mailto:liviu.o.pop@gmail.com)

### **Building Memory Boxes: An Open Source Approach**

#### **Keywords:**

Memory Boxes / Time Capsules / Raspberry Pi

This workshop attempts to teach non-technical participants how to create a memory box that can play sounds using a Raspberry PI, an RFID sensor and RFID tags. The box as a prototype was designed as a tool for returning songs and poems from folklore archives back to the villages where were recorded, a two or three generations later. In order to have a stronger connection with those memory boxes, we planned to create custom enclosures for them made from traditional materials, and in this particular case, from wood. We added small wooden coins with RFID tags that act as interfaces to the small memory boxes, thus lowering and hiding the technical aspects and making them more friendly and more familiar. We aimed to obtain a hidden tech interactive installation that will be easy to use. By using small wooden coins that have RFID tags embed (similar to the access cards used in automatically opening the doors), we hope to achieve a non-intrusive, friendly and familiar object instead of an alien hightech totem that nobody would approach and use.

Learn the Command Line

<https://www.codecademy.com/learn/learn-the-command-line>

Raspberry Pi Based Rfid Music Robot

<http://www.instructables.com/id/Raspberry-Pi-based-RFID-Music-Robot/>

Raspberry Pi Zero + OpenWrt, via USB On-The-Go

<https://medium.com/@matthewhuie/connecting-a-raspberry-pi-zero-to-a-network-via-usb-on-the-go-and-openwrt-8995a58ec151>

The Rosetta Disk

<http://rosettaproject.org>

The 10 Coolest Time Capsules Opened in 2015

<https://paleofuture.gizmodo.com/the-10-best-time-capsules-opened-in-2015-1747956945>

## Sarah Tytler

University of Glasgow,  
Glasgow, Scotland, United Kingdom

e-mail: s.tytler.1@research.gla.ac.uk

### **It Came from the Future! Integrating Digital Tools and Social Media into the Classroom**

#### **Keywords:**

**Inclusive Education / Multimodality /  
Research Practice**

Research and higher-educational academic practices are threatened by ideologies and policies which seek to reduce it to a linear consumer-production enterprise. Due to the increased rate of technological advancements, universities are often burdened with a seemingly impossible task: to prepare students for careers that don't yet exist. We believe that teachers should integrate technology in ways that enhance student engagement, increase access, and promote inclusivity while also supporting effective learning, critical thinking, and problem-solving in the face of the complex issues challenging our world.

We advocate the incorporation of digital creative writing tools as a way to bypass barriers to inclusivity and enhance curricula of any structure, in any discipline. In our interactive, multimodal, three-hour creative workshop, we will discuss and model different ways to use digital tools for creative writing and research to engage students, as well as provide participants with a fun, creative environment in which they can explore ways to integrate these tools into their own research and teaching practices. We recommend tools that support student-centred pedagogies and help us as scholars to envision our own work in new ways, focusing on creating inclusive academic environments and using digital and social media tools to engage both students and colleagues.

This workshop is for researchers, educators, and course designers who wish to engage with their practice in new ways and network with others who share their passion for innovative ideas about improving education for diverse learners. In addition to having a multimodal making-space, this workshop will include round-table and small-group discussions as well as networking opportunities that we hope will encourage exciting new collaborative possibilities.

We will first demonstrate a few techniques and discuss the pedagogical theory behind multimodality and creativity's role, as part of an interdisciplinary educational approach, in increased student engagement, participation, and knowledge retention. Then the workshop becomes a supportive, hands-on environment where participants will explore the benefits of multimodality providing new insights into their research practice through a series of guided, interactive stepping-stone writing activities and fun, multimodal exercises. By combining different types of learning--such as kinaesthetic, auditory, visual, and verbal--participants in this workshop will experiment with new ways teaching creative writing and writing for research by incorporating digital creative writing tools and social media into their classrooms.

By the end of this workshop, participants will be able to:

- Consider new approaches to research practices with customised activities;
- Access resources and research that supports integration of multimodal learning techniques and creative digital tools;
- Use examples and exercises from the workshop to share with colleagues at their home institutions;
- Consider how creative digital tools can be used to better address issues of inclusivity;
- Connect with professional networks that foster support and encourage collaborative projects.