

eite.dh

# DH\_BUDAPEST\_2018

Centre for Digital Humanities – Eötvös Loránd University



## Contents

### **Béla Adamik:**

Demonstration of the Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age

### **Zsolt Almási:**

Fitness for or Fitness of purpose?: The Database as Service Building the Hungarian Shakespeare Archive

### **János Bárh M.:**

The EHA-project: Database of Transylvanian Historical Place Names and Interactive Linguistic Map

### **Dávid Bartus – Zoltán Czajlik – László Rupnik:**

Born digital from ruins: preserving and disseminating digital cultural heritage. A case study of an aerial archaeological archive

### **Ágoston Zénó Bernád – Maximilian Kaiser – Katalin Lejtovicz – Peter A. Rumpolt – Matthias Schlögl:**

Corpus Analysis and Source Criticism – Measuring the Austrian Biographical Dictionary 1815–1950 (ÖBL)

### **Bernadett Csurgó – Judit Gárdos – Szabina Kerényi – Éva Kovács – Andras Micsik:**

Building, analysing. The COURAGE Registry of cultural heritage collections – empirical and epistemological analyses

### **Jānis Daugavietis – Eva Eglāja Kristsons:**

Creative Crowdsourcing for Heritage actualisation: analysis of campaign "Recite Veidenbaums' Poetry!"

### **Marek Debnár:**

DR2 methods in philosophical research

### **Laura Dietz – Claire Warwick – Samantha Rayner:**

Owning the 'Unreal': keeping and collecting digital novels

### **Bálint Dömölki:**

Structure of a historical data archive

### **Patrick Egan:**

Re-imagining Music Projects from the Seán Ó Riada Collection

### **Bettina Fabos – Leslie Waters – Kristina Poznan:**

Re-writing Hungarian history with an interactive website: the digital humanities timeline project Proud and Torn

### **Federica Fantone – Hugo Manguinhas – Valentine Charles – Antoine Isaac:**

Semantic enrichment in Europeana: a brief excursus across challenges and opportunities

5

6

7

8

9

10

11

13

14

15

17

18

19

**Anne Ferger – Daniel Jettka – Timm Lehmborg:**

Categorization of Language Documentation Data –  
A Graph-based Approach

20

**Daniel L. Golden:**

Epistemological commitments in digital humanities

22

**Andrea Hřčková**

Increasing the findability of digital heritage documents  
by using Search Engine Optimization methods

23

**Dimitar Illiev – Dobromir Dobrev – Grigor Boykov:**

Researching inscriptions and archives in the DH Lab  
to the University of Sofia: challenges, methods and perspectives

24

**Emese Ilyefalvi:**

Transdisciplinarity in digital scholarly editing?  
The perspective of a folklorist

25

**Tamás Kiss:**

Introducing Ottoman Turkish Text Analysis Software Rumi 1.0:  
A Quantitative Analysis of Gelibolulu Mustafa 'Alī's Works

26

**Tibor Koltay:**

Digital humanities: opportunities and challenges for librarians,  
library and information science and information literacy

27

**Alexander König – Verena Lyding – Elisa Gorgaini:**

Building a digital infrastructure in South Tyrol

28

**Jessie Labov – Anton Mudrak:**

Encrypted Channels, Distributed Networks: The Telex footprint  
in the Cold War and its legacy for media practices today

29

**Davor Lauc – Darko Vitek:**

From the History to the Story – Harvesting Non-Monotonic Logic  
and Deep Learning to Generate Multilingual Family Narratives  
From Genealogical Data

30

**Daniel McDonald – Eveline Wandl-Vogt – Mahsa Vafaie**

Integrating crowdsourcing and the blockchain for natural language  
data collection: novel methods, use-cases and debates

31

**Berrie van der Molen:**

The public framing of MDMA in Dutch cross-media debates. Developing  
a digital tool to answer historical research questions based on distant  
and close reading of cultural heritage “big data”

32

**Majlinda Muka – Dorina Xheraj-Subashi:**

The influence of digital storytelling on the achievement competences  
of students in cultural heritage teaching

33

**Anna Neovesky – Frederic von Vlahovits:**

IncipitSearch: A common interface for searching in music repositories

34

**Thomas Palfinger – Alexander Schatek –  
Amelie Dorn – Eveline Wandl-Vogt:**

Food cultures: co-creation and evaluation of a thesaurus  
as a cultural infrastructure

35

**Ioulia Pentazou – Ioanna Laliotou:**

Digitizing the History of Academia: Creating the Digital Archive  
of the University of Thessaly

36

**Róbert Péter:**

Metadata visualization of 18th-century British newspaper  
and periodical articles concerning Freemasonry

37

**Roel Smeets – Maartje Weenink:**

Opposite-Sex Relations in Present-Day Dutch Literature.  
A Network Analytical Approach to Character Representations

38

**Filomena Sousa:**

What about Digital Platforms of Intangible Cultural Heritage?

39

**Botond Szemes:**

The practice of the close reading, and its possible alternatives

40

**Miklós Tamási – András Török – Bettina Fabos:**

Duplicating the Fortepan Photo Archive in Neighboring Countries

41

**Bogata Timar – Erika Asztalos – Ditta Szabó – Nikolett F. Gulyas:**

The Typological Database of the Ugric Languages: establishing  
a long-term project

42

**Yun-Cheng Tsai – Pu Yu – Chia-Ching Wu – Ji-Yu Huang:**

Natural Language Processing Real-Time System for Central European News

43

**Lucia Vannini:**

Crowdsourcing projects in Classics: a reflection on models  
of collaborative editing of primary sources

44

**Fruzsina S. Vargha:**

Estimated acoustic parameters from digitised dialect atlases

45

**Melinda Vasari:**

Securing the Literary Evidence. Some Perspectives on Digital Forensics

46

**Cristina Vertan:**

Multilevel Annotation of Historical Documents

47

**Sara Wagner:**

Cultural heritage, digital access: the DIY archiving process  
of the Syrius band (1969–1973) and what it reveals

48

**Eveline Wandl-Vogt:**

Open Innovation Research Infrastructure:  
Value driven organisational designs fostering innovative DH  
on the example of Biographical research

49

## Béla Adamik

Eötvös Loránd University,  
Budapest, Hungary

e-mail: [adamik.bela@btk.elte.hu](mailto:adamik.bela@btk.elte.hu)

### **Demonstration of the Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age**

#### **Keywords:**

Digital Latin dialectology / Database of Latin Inscriptions / Linguistic geography

The aim of the project “Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age” is to develop and digitally publish (at <http://lldb.elte.hu/>) a comprehensive, computerized historical linguistic database that contains and manages the Vulgar Latin material of the Latin inscriptions found in the territory of the Roman Empire. This will allow for a more thorough study of the regional changes and differentiation of the Latin language of the Imperial Age in a wider sense and for a multilayer visualization of the discovered structures concerning linguistic geography as well as for a better understanding of the processes that led to the development of Romance languages and determined the linguistic, ethnic and even cultural features of medieval and modern Europe. The project is going to be realized with the collaboration of the Latin Department of the Eötvös Loránd University, Budapest and the Lendület (‘Momentum’) Research Group for Computational Latin Dialectology of the Research Institute for Linguistics of the Hungarian Academy of Sciences, supported by the National Research, Development and Innovation Office NKFIH (former Hungarian Scientific Research Fund OTKA; no. K 124170, 2018-2021, K 108399, 2014-2017; no. K 81864, 2010-2013 and no. K 62032, 2006-2009) and by the ‘Momentum’ Program of the Hungarian Academy of Sciences (2015-2020) (cf. <http://lldb.elte.hu/>).

The proposed paper serves as an introductory presentation of the Database considered here, which is first of all intended for a wider but professional public interested in a project that deals with the dialectology of the Latin language concerning Roman inscriptions, according to a special linguistic approach, with the help of modern information technology.

## Zsolt Almási

Péter Pázmány Catholic University,  
Budapest, Hungary

e-mail: [almasi.zsolt@btk.ppke.hu](mailto:almasi.zsolt@btk.ppke.hu)

### **Fitness for or Fitness of purpose?: The Database as Service Building the Hungarian Shakespeare Archive**

#### **Keywords:**

[Online database](#) / [Cultural heritage](#) /  
[Digital edition](#) / [Digital Archive](#)

This paper aims to elucidate the problematics lying in the possible opposition between principles of the Fitness for the Purpose and the Fitness of the Purpose when assessing/building an online database, resource for (digital) scholarship. The former principle prescribes deploying the most cutting-edge technology, the inclusion of the greatest amount of content—whatever that may be depending on the theme of the database. While prescribing these, this principle also implies that the building of the database will be in need of resources financial and human at the same time, which in turn entails that the building of the infrastructure will be also time-consuming as well. A negative effect of this principle will inevitably lead to the widening of the “digital divide.” The negative effects of this principle can be balanced if the latter principle is considered as well, which lies in exploring the needs of the target audience, the future users who may not need the most advanced technology, the most expensive solutions, but need a resource that can be used to solve the most immediate problems the community faces. The second principle, while seemingly going against the first one, may well be seen as a principle to moderate the prescriptions of the first one, so as to force the builders of the database to create an infrastructure that will be beneficial for the target users without necessarily giving up the dreams of cutting-edge technology. The two principles together in the right balance define then an online database as a specific type of service for the given scholarly community.

The definition of a database as service will be exemplified with a reference to the Hungarian Shakespeare Archive. The Hungarian Shakespeare Archive is in its beta phase at the moment, but even in this state, it is possible to show what may lie in the concept of service for the target community. Although the building of the database is full with compromises following from the lack of appropriate financial backing, from the constraints of the CMS, from the problematics of completeness in so far as the Shakespearean textual condition is concerned, yet even at this early phase it may well automate tasks Shakespeare scholars face while carrying out their research and writing up reports on their findings and also it should help clarify research questions and pose new ones as well.

## János Bárth M.

Eötvös Loránd University,  
Budapest, Hungary

e-mail: [janos.barth@gmail.com](mailto:janos.barth@gmail.com)

### **The EHA-project: Database of Transylvanian Historical Place Names and Interactive Linguistic Map**

#### **Keywords:**

[Place Name / Database / Dialects /  
Attila Szabó T. / Transylvania / map](#)

This paper presents a website that publishes Attila Szabó T.'s collection of Transylvanian historical place names (Szabó T. Attila Erdélyi történeti helynévgyűjtése) in a searchable, online format. In this database the place names are linked to the coordinates of the given settlement, thus they can be displayed on map as well.

The toponymical collection contains about 350 000 historical place names from Transylvania in the 14th-20th centuries, which were collected from all kinds of written sources. The place name dictionary was published between 2001-2010 based on Attila T. Szabó's manuscript collection. The collection consists of eleven volumes, arranged according to the former Transylvanian comitats. The online version will hopefully serve as an up-to-date index, search engine for both Hungarian and international Onomastics and also for historical and social scientific research.

There are several searching options in the database: with the help of the search engine, any letter combination, word, name or name structure can be found, together with its chronological, settlement, type attributes and its context. With the help of the coordinates, one can produce a map showing the locations of occurrences of the given linguistic form. For this purpose, the popular and most user-friendly Google Maps has been chosen. These maps are capable of showing us a number of interesting regional linguistic differences through historical name data.

This makes it a major tool for onomastics – researching territorial differences in name giving/ name usage, historical linguistics and dialectology, but historical science and archeology can also make great use of it. In my lecture, I'll show examples where spectacular regional differences can be explained by social and historical phenomena.

**Dávid  
Bartus –  
Zoltán  
Czajlik –  
László  
Rupnik**

Eötvös Loránd University,  
Budapest, Hungary

e-mail:

bartusdavid@gmail.com

czajlik.zoltan@btk.elte.hu

rupniklaci@gmail.com

**Born digital from ruins: preserving and disseminating digital cultural heritage. A case study of an aerial archaeological archive**

**Keywords:**

Digital Cultural Heritage / Born Digital Documents / Online Database / Aerial Archaeology / Aerial Photography

Aerial archaeology is an airborne remote sensing method to identify and record archaeological information in a non-destructive way, mostly by searching for cropmarks, through which sub-surface archaeological features can be visible from the air. The main tool of aerial archaeology is taking pictures from an airplane, or by an unmanned aerial vehicle, commonly known as a drone. The results of the last two decades of aerial photography are born digital documents, moreover, since most of the identified sites are only visible from altitude in a very short period of the year, these photographs are often the one and only documents of an archaeological site.

The Institute of Archaeological Sciences (Eötvös Loránd University) has more than 65.000 photos taken by aerial archaeology. In a recently started project, we aimed to build an online database to preserve and disseminate these photographs with the relevant geospatial information on the cultural heritage which they represent. The present paper will discuss the methods, problems and results of the project.



**Ágoston  
Zénó Bernád –  
Maximilian  
Kaiser –  
Katalin  
Lejtovicz –  
Peter  
A. Rumpolt –  
Matthias  
Schlögl**

Austrian Academy of Sciences,  
Vienna, Austria

e-mail:

agoston.bernad@oeaw.ac.at  
maximilian.kaiser@oeaw.ac.at  
katalin.lejtovicz@oeaw.ac.at  
peter.rumpolt@oeaw.ac.at  
matthias.schloegl@oeaw.ac.at

**Corpus Analysis and Source Criticism  
– Measuring the Austrian Biographical  
Dictionary 1815–1950 (ÖBL)**

**Keywords:**

Digital Humanities / Biography / Austria-  
Hungary / Corpus Analysis / Biographical  
Dictionary

Quantitative evaluations in the humanities have become more and more important in the recent years (e.g. Haber 2011), and the term “Big Data” is also being frequently used in connection with humanities in general (e.g. gral 2016). Analyses of textual resources are often based on uniform corpora (e.g., Wikipedia: Russo et al., 2015), which does not only offer a standardized way to access thousands of records, but also to develop and take advantage of homogeneous metadata and harmonized vocabularies. While these corpora are well suited for analysis, the underlying data is often unbalanced, as the methodology of data collection can lead to an incomplete or irregular information pool. Heterogeneity of data especially affects national biographies, where the material is collected and compiled by different authors over a long period of time. Despite the dilemma of distorted datasets, studies tend to focus on the reliability of the digital tools themselves rather than the reliability of the resources (e.g. Reinert et al. 2015; Stotz et al. 2015).

The presentation offers an exemplary corpus analysis of the digital version of the Austrian Biographical Dictionary 1815-1950 (ÖBL). Founded in 1946 and published since 1954 (since 2009 online), the ÖBL covers not only the territory of present-day Austria, but the entire Habsburg empire, thus providing an image of Central European culture between 1815 and 1950 (Obermayer-Marnach 1957; Lebensaft, Reitterer 1996; Gruber, Feigl 2009). The not yet completed reference work contains about 20,000 biographies in 14 volumes (68 issues). In this presentation we discuss the often neglected inconsistency of a seemingly unified resource. The analysis demonstrates the variations of the biographies in the ÖBL in regard to information density (number of Named Entities relative to the number of tokens in the biography), countries of origin (the geographical distribution of birthplaces) and gender (proportion of female personalities depicted in the ÖBL over the more than seventy-year history of the lexicon). By discussing this and related topics, we would like to address which challenges appear by using national biographies as a resource for answering quantitative historical research questions.

**Bernadett  
Csurgó –  
Judit  
Gárdos –  
Szabina  
Kerényi –  
Éva  
Kovács –  
Andras  
Micsik**

Hungarian Academy of Sciences, Center for  
Social Sciences, Budapest, Hungary  
SZTAKI Budapest, Hungary

e-mail:

csurgo.bernadett@tk.mta.hu

gardos.judit@tk.mta.hu

kerenyi.szabina@tk.mta.hu

kovacs.eva@tk.mta.hu

micsik@sztaki.hu

**Building, analysing. The COURAGE Registry of  
cultural heritage collections – empirical and  
epistemological analyses**

**Keywords:**

Archive / Digital History / Sociology /

Opposition / Socialism / Culture

COURAGE (“Cultural Opposition – Understanding the CultuRal HeritAGE of Dissent in the Former Socialist Countries”) is a three-year international research project funded by Horizon 2020, the EU Framework Programme for Research and Innovation.

COURAGE creates a comprehensive online database (digital registry) of existing but scattered collections on the histories and forms of cultural opposition in the former socialist countries and thereby makes them more accessible. It analyses these collections in their broader social, political and cultural contexts. The general aim of this analysis is to allow for the expanded outreach and increased impact of the collections by assessing the historical origins and legacies of various forms of cultural opposition. COURAGE creates a registry of collections that relate to various forms of cultural opposition in a vast geographical area embracing all member and potential member states of the EU in the former socialist bloc. The Registry provides standardized descriptions of the collections. COURAGE includes research on the social and cultural practices generated by the collections, which themselves represent measures of faith in the potential of the archive as a repository of the past. In our paper, we analyse the types of collections that can be found in the Registry about the collections of opposition in socialism: our initial aim was to include not only the most known ones, but also small collections on the periphery of the archival world. We explain how our own research questions shaped the online Registry and the methodology behind it, especially the interview guideline that underlines the work of the researchers filling in the Registry. We discuss how our international and interdisciplinary research groups shaped the process of the making of the Registry and its usage.

We describe how the public surface presenting the Registry, which has been selected from the entirety of the metadata, and explain how the Registry itself functions. We draw some conclusions regarding how the ontology of such a data structure can influence and/or generate scientific practices. We try to summarize how the different types of data in the Registry enable different types of scientific conclusions and/or definitions of cultural dissent. The paper also provides general statistics about the collections, their typology, locations, age, year of fundings, etc. We also analyse the dynamics of the registry content, the growth of data, amount of problems found, typical mistakes, correction time, etc. We aim to give a general overview of the field of research and provide the readers with the first results of the analysis of the data in the Registry.

# Jānis Daugavietis – Eva Eglāja Kristšone

University of Latvia,  
Riga, Latvia

e-mail:

janis.daugavietis@gmail.com

kristšone@gmail.com

**Creative Crowdsourcing for Heritage  
actualisation: analysis of campaign “Recite  
Veidenbaums’ Poetry!”**

**Keywords:**

Creative Crowdsourcing / Digital Cultural  
Heritage / Digital Reflexivity / Digital  
Participation / Latvia

This paper seeks to analyse the process and results of innovative creative crowdsourcing campaign carried out by the Institute of Literature, Folklore and Art (University of Latvia) in 2017 in which people were asked to recite, record and submit the poems of Latvian poet Veidenbaums using specially programmed crowdsourcing tool. Afterwards, using online questionnaire, crowdsourcers were surveyed for their socio-demographic characteristics, motivation to participate and evaluation of the different aspects of the campaign from technical ones to personal ones.

From September to December 2017 as a part of 150th anniversary of a Latvian poet Eduards Veidenbaums’ Latvian society was invited to virtually engage to participate in campaign by reading one or more of his poems aloud and recording. Recording was made possible by using everyone’s computer or mobile phone and/or uploading it to the site <http://lasi.literatura.lv>. Another opportunity to participate in campaign was provided by the specially built ‘Veidenbaums Studio’ — a mini-recording studio situated at The Latvian National Library.

The overall theoretical framework of the project behind this campaign is “digital reflexivity, aiming at the understanding of specifics, conditions and consequences of participatory knowledge production in the digital age.” (‘Empowering knowledge society: interdisciplinary perspectives on public involvement in the production of digital cultural heritage’, project proposal). In order to analyse process and results of aforementioned campaign, the following objectives are put forward:

- evaluation of the designed tool, focusing on the process of involvement and communication;
- analysis of submitted poetry readings (audio records);
- analysis of the social groups who participated (and left outside).

This was the first pilot attempt within the broader project to create the user-oriented digital resource which could be widely usable in the system of education – from primary to higher level, as well as in lifelong learning. Therefore, the analysis of the process of “Recite Veidenbaums’ poetry” has practical meaning. Analysis of other results (eg. poems recited by people, collected date of crowdsourcers and their answers to the survey questions) allows to make more theoretical justification of the project. Recitation of a literary texts as a part of crowdsourcing through online activities asks for a subtler analysis and proposes a cultural perspective on the “reflexive relationship between user and medium” as a “remediation” of the self (Bolter and Grusin, 2000: 230).

A conceptual (and also political) problem is fostering accessible digital participation. Within the project one of our aims is to promote social inclusion and support identity-building and value-defining processes of the society, providing a participatory digital environment. Analysing data from current

campaign, we will try to understand the level of a non-discriminatory public involvement in creation of digital cultural heritage.

Crowdsourcing campaign “Recite Veidenbaums’ poetry” turned out to be a successful connection with cultural heritage and powerful platform for individual engagement. Irrespective of final conclusions, our preliminary analysis of the campaign suggests that the creative and performing arts have potential as an application in areas of education and life learning also through digital tools.

## Marek Debnár

Constantine the Philosopher University,  
Nitra, Slovakia

e-mail: mdebnar@ukf.sk

### **DR2 methods in philosophical research**

#### **Keywords:**

Digital Humanities / Philosophy / Distant  
Reading / Franco Moretti / Quantitative  
Formalism

Distant reading and data-driven research (DR2) is an umbrella term for broad spectrum of approaches and methods of quantitative research within digital humanities. The use of DR2 approach in philosophy builds on principle of distant reading (a term by Franco Moretti, currently used for various methods in IT processing of texts and bigger corpora), as well as on principle of quantitative formalism. Nowadays, this approach is being explored in various projects at several leading centers for philosophical research (like DR2 group - Distant reading and data-driven research in the history of philosophy at University of Torino). The study aims to describe and analyze DR2 methodology with regard to contemporary inquiries, and to define specific philosophical features of DR2 in comparison to literary science, from which this method originated and where it had been applied. The study seeks to answer question, whether it is possible for DR2 research to discover hidden or overseen premises and arguments in philosophical texts, what are the requirements and specific features of philosophical text corpora with regard to research objectives (level of annotation, etc.), what are the challenges with respect to development and usage of already accessible IT tools in philosophical research.

Laura  
Dietz -  
Claire  
Warwick -  
Samantha  
Rayner

**Owning the 'Unreal': keeping and collecting digital novels**

**Keywords:**

E-Books / Reading / Ownership / Personal Library / Born Digital

This 20 minute individual presentation will introduce results from a current research project on the reputation and legitimacy of digital novels, focussing on the specific question of how and why readers keep and collect digital novels, particularly born digital works. It will draw on recently collected original data on e-reading from a survey (981 respondents) and linked focus groups and interviews (40 participants) to examine how the ways in which digital novels can and can't be held (or discarded) interact with understanding of their status and 'realness'. It will present findings on attitudes towards conditional use licenses and piracy, the role of Amazon's image and business practices, and reader strategies for obtaining books legally and illegally.

Findings suggest an increasingly complex environment for e-novel and more general e-book ownership, to which readers respond with increasingly nuanced and flexible conceptions of 'keeping' and 'giving away'. A decade after Kindle, long-term relationships to non-physical books are no longer hypothetical, and readers continue to develop their requirements for both utility and satisfaction. Print remains preferred in almost every circumstance, but a significant minority of participants now choose digital as better for keeping as part of a personal library. E-books, and e-novels in particular, still serve as useful 'fakes' (Dietz, Warwick and Rayner, 2015: 26-7) for many readers in many situations, but emerge – when their readers need them to be – as sometimes-legitimate real books, with implications for the future of personal libraries as well as the future of novels. Control and agency, more than 'realness', are the essential components for meaningful ownership of a digital book. The threat of books being lost, or repossessed by retailers, is a particular concern; a fear of such actions is enough to drive many readers away from digital (as a blanket policy or as the format for a given book). The sense of meaningful ownership can, however, be taken back, via principled resistance, 'digital audition' (Dietz, Warwick and Rayner, 2015: 30) , or a conscious decision to change one's concept of what an e-book can be and accept a digital book collection as a genuine personal library. This recasts e-books as an integral part of building a personal library: sometimes as components, but sometimes just as tools. This mixed usage, combining conceptions of e-novels as real books, useful fake books, parts of books, and digital proxies, further demonstrates how readers are able to move between conceptions as the situation demands. The paper's theoretical framework will draw on Publishing Studies and Book History as well as Digital Humanities, and will engage with Price and de Certeau on reader strategies in the face of institutional control, Schwabach on readers' responses to copyright restrictions, and Buchanan, McKay and Levitt on reading device usage. With its focus on ownership and legitimacy of digital-original e-novels, this paper will align with conference topics, most particularly the thread on digital cultural heritage and born digital documents.

## Bálint Dömölki

John von Neumann Computer Society,  
Budapest, Hungary

e-mail: bdomolki@gmail.com

### Structure of a historical data archive

#### Keywords:

Data archive / Information Technology /  
History / Hungary

Collection and distribution of factual data about the history of different fields of technology can be regarded as an important part of the preservation of our cultural heritage. Data archives can be built to perform this task for a specific subject, geographical area and time period.

A generic model for such archives might be defined with the following properties:

- Collecting all documents (text, image, video, voice, ...) in a repository;
- Defining sections to describe specific aspects of the given field (e.g. persons, institutions, products etc.) with composed descriptions of the individual items, containing references to the documents and to each other;
- Items in each section can be viewed in the form of data-sheets, presenting all information about the item and also in the form of tables, listing one item per row, with the most important properties;
- Allowing the widespread use of cross-referencing within the archive and also to objects of the outside world;
- Including a sophisticated search system for finding individual items or lists of items with common properties;
- Providing tools for creating and editing the archive, with the possibility of parallel building of (inter-related) sections;
- Organizing interaction with the users, allowing them to send feedbacks, including proposals for modification/addition.

The model is demonstrated on the example of a data archive for the history of the Information Technology field in Hungary, from the 50s to present days, with emphasis on the years before 1990. The Archive is created by the IT History Forum (iTF) of the John von Neumann Computer Society (NJSZT). It can be reached at [itf2.njszt.hu](http://itf2.njszt.hu) (presently in Hungarian, English version is being prepared). Implementation is done in WordPress.

The documents being collected are

- Writings (books, publications, manuscripts etc.) related to Hungarian IT history,
- Videos, including oral history interviews and recording of talks at iTF events etc.,
- Historic photos and picture galleries.

Other Sections include

- Persons, having some relation to IT (both alive and deceased),
- Institutions (in business, research, education, administration etc.) having been active in IT,

- Products (hardware, software) developed and/or intensively used in Hungary,
- Events (including conferences, seminars, workshops, contests etc.) organized in Hungary in the past 60 years.

The IT History Archive at present (Jan. 2018) contains over 2500 items, dynamically increasing. A few specific problems having been solved during the implementation:

- Due to the parallel development of the sections, it is possible to refer to an item not existing yet in that section
- Names of the institutions may change time to time, but we have to use everywhere the “historical” names.

The Archive is being developed by a team of 15-20 volunteers, in close cooperation with the IT History Exhibition of the NJSZT, located in the Szent-Györgyi Albert Agora, Szeged, collecting the physical objects related to IT history in Hungary.

The structure of the Archive and its software solutions might be adapted to the creation of similar historical data archives (or their parts) for other subject areas.



# Patrick Egan

University College Cork,  
Cork, Ireland

e-mail: [patrick.egan@student.ucc.ie](mailto:patrick.egan@student.ucc.ie)

## Re-imagining Music Projects from the Seán Ó Riada Collection

### Keywords:

[Digital Humanities](#) / [Ethnomusicology](#) /  
[Metadata](#) / [Practice Theory](#)

This paper attempts to critique novel digital practices utilised alongside research in ethnomusicology to account for the rise in popularity of a twentieth century artist in Irish society.

The exploration examines evidence from the music-related projects of Seán Ó Riada (1931-1971), a popular Irish artist whose creative output became a major contributor to notions of Irish expression during his lifetime.

I begin with an overview of the Seán Ó Riada Collection (SÓRC) held at the Boole library, University College Cork. Archival practices are outlined in order to uncover key information relating to material acquisition, archiving processes and standards which facilitated the creation of a thematic collection.

ata is central to the contextualisation of historical events, through focusing on descriptions of the circumstances through which documents from within the SÓRC were performed (in the broadest sense). Computer coding is then used to create digital representations of the materials and their descriptions. The work proceeds with the hypothesis that using close reading of archival items and representing them within digital prototypes, we may ask more informed questions about the materials in a special collection, and move our engagement with the material beyond traditional thematic description.

In tandem with the analysis and digital representation of the collection materials, I reflect upon qualitative and quantitative methods of data gathering. This includes external archival visits, digitised newspaper searches and tags within literature concerning Seán Ó Riada. Following a digital re-representation of this combined data, practice theory is advanced as a suitable methodology for historical analysis. Key informants are selected and interviewed from a wide range of backgrounds in order to contextualise historical circumstances, trends and principles of musical expression within Irish society. This theoretical approach facilitates a mixed methods investigation of musical practice in order to reveal evidence of change or continuity during the development of the artist's career.

The paper explores how the use of metadata with digital prototypes relates to explorations of research in ethnomusicology. This approach considers if digital humanities modelling and processes of digital research may inform traditional research practices in meaningful ways.

**Bettina  
Fabos –  
Leslie  
Waters –  
Kristina  
Poznan**

University of Northern Iowa, Cedar Falls,  
Iowa, United States

Randolph-Macon College, Ashland,  
Virginia, United States

Journal of Austrian-American History, Penn State  
University, Pennsylvania, United States

e-mail:

fabos@uni.edu

lesliewaters@gmail.com

kepoznan@gmail.com

**Re-writing Hungarian history with an  
interactive website: the digital humanities  
timeline project Proud and Torn**

**Keywords:**

[Hungary](#) / [Interactive](#) / [Parallax](#) / [History](#) /  
[Archival Photographs](#)

The interactive digital history website, Proud and Torn: How My Family Survived Hungary, ([proudandtorn.com](http://proudandtorn.com)) is a unique and groundbreaking public history and digital humanities project that will greatly advance our understanding of what it means to be Hungarian. The animated, digital timeline is four years in the making and an ambitious historical narrative adapted to the web. The timeline contains close to a thousand images from 37 different photographic archives and links to 16 drop-down chapters that each introduce a particular period of Hungarian history. The timeline itself and each individual chapter stylistically combine the genres of photomontage and graphic history and present the content with parallax scrolling, a special web coding technique that makes background images move slower than foreground images, creating an illusion of depth and a more immersive visual experience.

The Proud and Torn web project tells the story of an everyday Hungarian family from the countryside, 1840-1956. Traditional Hungarian histories are about public figures and life in and around Budapest. They typically emphasize Hungary's "great states and statesmen," focus on nationalism and national mythmaking, and extol the lives of the privileged and the powerful. This is the history portrayed by the National Museum in Budapest: an endless parade of noble Magyar men. These prominent figures, and the historical discourses that perpetuate their significance, dominate the way we understand Hungarian history.

As an alternative history, Proud and Torn emphasizes the lives of everyday Hungarians, focusing on one family and the people they knew and did business with in a small Hungarian farming town near Lake Balaton. We limited our use of the most conspicuous and iconic photographs because these images, compelling as they may be, are ultimately limiting our interpretation of history. We've relied instead on alternative visual sources to document rural Hungary, including this one Hungarian family's own collection of photographs and the photographic collections of the Fortepan digital archive and the Hungarian Police Photo Archive.

Fortepan ([fortepan.hu](http://fortepan.hu)) is a public treasure, with a deliberate mission to draw attention to the subtle and lyrical work of anonymous amateur photographers throughout Hungary and to bring the images of everyday, non-public existence (1900 – 1990) to light. The Hungarian Police Photo Archive documents the history of communist Hungary through the viewpoint of the Ministry of Interior—those responsible for carrying out the plans of collectivization, interrogation, and the maintenance of a communist state. In effect, the now-digitized collection of photographs document best practices on "how to be the perfect socialist policeman," and thus contains incredible images of policemen raiding family farms and arresting land-owning farmers—both women and men. Both of these archives are supported by the Open Society Archive in Budapest. The Proud and Torn project is a testament to their value and significance.

In the end, Proud and Torn expands our idea of Hungary by placing a greater emphasis on rural and agricultural history and using fresh visual sources from amateur and underutilized archival collections. It is an important contribution to digital humanities in Central Europe.

**Federica  
Fantone –  
Hugo  
Manguinhas –  
Valentine  
Charles –  
Antoine  
Isaac**

Europeana Foundation, The Hague, Netherlands

e-mail:

[federica.fantone@europeana.eu](mailto:federica.fantone@europeana.eu)

[hugo.manguinhas@europeana.eu](mailto:hugo.manguinhas@europeana.eu)

[valentine.charles@europeana.eu](mailto:valentine.charles@europeana.eu)

[antoine.isaac@europeana.eu](mailto:antoine.isaac@europeana.eu)

### **Semantic enrichment in Europeana: a brief excursus across challenges and opportunities**

#### **Keywords:**

[Automatic Metadata Enrichment / Alignment  
of Vocabularies / Geonames / Dbpedia and  
Wikidata / Europeana Entity Collection /  
Linked Data Strategy](#)

Europeana faces many challenges dealing with the heterogeneity of data. This is due to a variety of factors including the large amount of objects and partner organizations, the diversity of content, the large number of represented languages, the variety of quality level, the multiplicity of rights statements, the high number of contextual references. The implementation of the Europeana Data Model has allowed for more flexibility to metadata granularity and interoperability, for support of contextual resources and for reliance on (Linked Data based) vocabularies for contextualisation. The adoption of the Linked Data approach has already introduced some changes. But the improvement of user services in terms of semantic richness, multilingual coverage and re-usability remains crucial. There is no unique solution; the integration in terms of both tools and vocabularies offers the opportunity to face those challenges and to benefit from shared opportunities. Seen as a platform, Europeana re-uses vocabularies and datasets from third-party sources for contextualisation and performs automatic enrichment based on them. In addition, Europeana promotes the contribution of vocabularies from data providers and encourages the alignment between them. Europeana has included Geonames, Dbpedia and Wikidata in its enrichment strategy. This choice has been motivated by the benefits these extensive knowledge bases provide in terms of, for example, multilingual coverage, variety of information, consistency in language labelling, co-reference to other vocabularies. For example, Dbpedia provides co-reference links to large datasets such as Wikidata and Freebase. Wikidata can support annotation in Europeana and, in turn, can benefit from enriching its knowledge base with GLAM data. But for Europeana's needs the use of only pivot vocabularies is not enough. Europeana has developed its own knowledge graph, the Europeana Entity Collection, which is meant as reference dataset collecting metadata for Europeana's entities. The Europeana Entity Collection will improve users' experience in their search and navigation through a network of connected resources. But choices have to be made. On this respect, the selection of data sources represents a crucial point. The EuropeanaTech Task Force on Evaluation and Enrichment has defined some criteria for the selection of potential target vocabularies or datasets. These are evaluated on the basis of their availability and access, granularity and coverage, quality, connectivity and size. Potential target vocabularies should complement each other; for the sake of consistency, it is important to find a balance between more pivot vocabularies covering large number of entities and more complementary vocabularies covering more domain specific entities. In order to fully disclose all the opportunities given by the re-use of metadata and content the engagement between Europeana, Wikidata and GLAMs is essential. In terms of definition of requirements and integration the discussion is open. For example, an Europeana ID property already exists in Wikidata as "identifier in europeana.eu for books, paintings, films, museum objects and archival records that have been digitised throughout Europe" and other Europeana identifiers in particular for entities of the Europeana Entity Collection are envisioned. This paper reviews and illustrates some of the steps Europeana has carried out so far in order to optimise the semantic enrichment with a focus on its Linked Data strategy.

Anne  
Ferber -  
Daniel  
Jettka -  
Timm  
Lehmberg

Akademie der Wissenschaften in Hamburg,  
Hamburg, Germany

e-mail:

anne.ferger@uni-hamburg.de

daniel.jettka@uni-hamburg.de

tim.lehmberg@uni-hamburg.de

### **Categorization of Language Documentation Data - A Graph-based Approach**

#### **Keywords:**

Graph Databases / Big Data / Language  
Documentation / Data Curation / Metadata

The contribution demonstrates a bottom-up approach to introduce the basic ideas and principles of graph-based data management to be used for the catalogization and curation of language documentation data in the framework of INEL (“Grammatical Descriptions, Corpora and Language Technology for Indigenous Northern Eurasian Languages“), a long-term research project in the context of the Academies’ Programme, which is coordinated by the Union of the German Academies of Sciences and Humanities .

The project aims at enhancing the possibilities for in-depth grammatical research on lesser documented (and in many cases endangered) indigenous languages on the territory of the Russian Federation. For this purpose (beyond other things) materials, including handwritten texts, analog sound recordings, word lists, etc., which are currently located in archives and collections (e.g. in Hamburg, Helsinki, Moscow, St. Petersburg, Tartu and Tomsk) are collected, curated and compiled to a comprehensive digital research infrastructure. As an integral part of this work detailed information on resources, subjects, researchers, etc. is collected together with geographical and temporal information where available with the aim to create a knowledge resource, which not only documents the presence of data and collections of data, but rather represents contents and relations between entities like research items, institutions, persons, and locations, and potentially reveals unexpected correlations in the data.

The initial step towards implementing a structured inventory of this data consisted in transforming existing descriptions of research data (texts, audio recordings in differing formats, transcriptions of spoken language, scientific publications, information about scientific tools, languages and researchers), which previously was mainly collected in MS Office files with distinct and partly concurrent descriptive categories. This requirement posed several basic challenges for the establishment of a comprehensive method to manage the information:

(1) relevant domains and categories of descriptions have to be identified and existing information has to be unified to ensure consistent data storage and management,

(2) a facility to continuously add consistent research data has to be provided and made available and usable for user groups with differing expertise in the project context.

(3) preparations have to be made to integrate new datatypes and descriptions for information that is to be collected in the future,

(4) the data should be searchable and be made publicly available.

After several steps of formalization and on the basis of the graph database Neo4j, a knowledge graph was built that covers and combines several domains of the collected research data, e.g. a bibliography, a tool inventory, and a general research data inventory.

Since in the initial phase of the project no comprehensive graphical user interface was available that allowed data curation and collection by project members, web-based frontends for managing and

editing the graph database content were implemented. At the same time intuitive graph-visualizations provided by existing JavaScript-based frameworks were used to harmonize and also analyse the data.

Based on this description and the experience made in the INEL project, the potential of graph databases for state-of-art approaches to lesser structured or complex data in the light of the special demands of language documentation will be discussed.

Thereby one focus will be on the compatibility of the data structures and vocabulary to be used with existing research data infrastructures and metadata catalogues like the CLARIN Virtual Language Observatory (VLO) or Open Language Archives Community (OLAC) .

## Daniel L. Golden

Hungarian Academy of Sciences,  
Institute of Philosophy,  
Budapest, Hungary

e-mail: golden.daniel@t-online.hu

### **Epistemological commitments in digital humanities**

#### **Keywords:**

Digital Epistemology / Data Models / Knowledge  
Representation

While digital humanities became one of the most intriguing and prosperous fields within the aging field of philology, both in terms of innovative power and constant funding, the very nature of scholarship has suffered substantive changes as well. The more the technological frameworks came into the front, the less the original conceptual schemes seem to succeed in feeding them with relevant and genuine knowledge. In my suggestion this can be traced back to the problem of methodology: while the tools and the phenomena of his investigations were clear and evident for the traditional philologist, his late digital colleague often has to grow himself into a philosopher of epistemology. Not only the objects of inquiry change from the analogue to the digital, but also all means accessing them. Arguably a whole new vocabulary for digital epistemology shall be developed to cope with the rapidly changing virtual reality. Generalizing the problem we can see digital humanities' scholars in a constant constraint of reinventing their epistemological toolbox in order to be able to build ever new and new ontologies of entities for their scholarly enterprises. One of the most striking examples of recent developments is the much acclaimed FRBR (Functional Requirements for Bibliographic Records) model, which was declared to make a revolution in document proceeding. However, a closer look to it will show its theoretical flaws and, let us call it that way, the naïve epistemology incarnated in it. This claim can be supported by a thorough analysis of the basic assumptions lying beneath the axiomatic concepts used by the model as a system for knowledge representation. A lot of intense work has been going on in the past three decades on data models, mark-up languages, text and character encoding standards etc. in order to produce better and better digital representations of our cultural heritage. The commands of preservation and distribution are seemingly unquestionable, therefore the quest for finding the ideal data structure is still thrilling. However, having already a historical corpus of that kind of projects, they can be subjected to theoretical reflection and analysis. In my presentation I plan to give a brief sketch of the epistemological adventures of the past and then try to draw some general conclusions about the possible directions digital humanities may want to take in this regard in the future.

## Andrea Hrková

Comenius University,  
Bratislava, Slovakia

e-mail: [hrckova.andrea@gmail.com](mailto:hrckova.andrea@gmail.com)

### **Increasing the findability of digital heritage documents by using Search Engine Optimization methods**

#### **Keywords:**

Findability / Search Engine Optimization / Digital Cultural Heritage

Mass digitization of cultural heritage and the growing amount of born digital heritage documents enabled the availability of these documents to everyone. Nevertheless, in the age of big data, the availability of content is not sufficient, if it means to be discovered and reached by the public. Nowadays, Google processes on average 40,000 search queries per second, therefore this medium cannot be ignored by any subject, striving to be discovered on the Internet. Still we face an underestimation in the field of findability of digital cultural heritage. The position of a content in search engine ranking page (SERP) might be influenced by the utilization of search engine optimization (SEO) methods that are rarely used in cultural sphere. These methods include among others the appropriate user research, semantic markup according to Schema vocabulary or microformats and sufficient number of links and mentions outside the website. In this contribution, the case studies of Slovak digital heritage portals using (consciously or unconsciously) SEO methods are presented and the possibilities for further development and research are suggested. The question is, whether these methods are able not just to increase the visibility of the documents of cultural institutions, but also to contribute to the decrease of disinformation and low quality content findability on the Internet by superseding this content from the first search engine result page.

**Dimitar  
Illiev -  
Dobromir  
Dobrev -  
Grigor  
Boykov**

Sofia University,  
Sofia, Bulgaria

e-mail:

dimitar.illiev@gmail.com

dobri@dobri.eu

griboykov@yahoo.com

**Researching inscriptions and archives in the  
DH Lab to the University of Sofia: challenges,  
methods and perspectives**

**Keywords:**

Ancient Greek Inscriptions / EpiDoc /  
Historical GIS / Digital Humanities / DH Lab

A DH Lab is currently being established to the University of Sofia as a result of the research activities of a cross-disciplinary group of scholars in several directions. A team of historians and philologists are working on the description and cataloguing of the rich archival collection of Slavic, Ottoman, Greek and other documents of the Bulgarian Zographou Monastery on Mount Athos. Epigraphers and classicists are preparing the publication of a corpus of Ancient Greek inscriptions from Bulgaria encoded in EpiDoc TEI. At the same time, a large group of researchers is finishing an Historical GIS model of the region of Thrace that includes different geo-localised objects across a great span of historical periods, using different types of sources as a basis. One of the things in common between all these initiatives, and one of the research topics that motivated scholars with different backgrounds and interests to find common methodological ground and eventually form the current DH Lab team, is the work on pre-modern original source items such as inscriptions on stone and archival documents.

The current presentation focuses on the peculiarities of such pieces of heritage and the particular problems that handling them poses with regard to the different approaches and purposes of the above-mentioned initiatives. First of all, such items lend themselves to description in at least two ways: as texts and as material objects with distinct physical features, provenance and history, as well as location and circumstances of discovery that may give additional information value to the texts contained in them. This is especially evident when the data from such documents is incorporated in a geospatial map model. Secondly, the information contained in such sources can be particularly useful for the enriching of the prosopography of a certain historical period and shedding new light of its topography and toponymy. Thus, extracting lists of people and places from the documents as well as establishing networks of relations between them is a task with multiple possible applications across various DH projects. Thirdly, the development of initiatives focusing on stone inscriptions and archives requires the recruitment of scholars with significant specialist skills in fields such as epigraphy, palaeography, etc., and their training to use DH instruments and methods, which is a task implying interesting challenges, logistic and methodological.

The presentation will provide examples of these and other issues, both common for inscriptions and archives and specific for each of them, and will reflect on the question how they influence the research and workflow at our DH Lab and the expected results of the different initiatives.



## Emese Ilyefalvi

Hungarian Academy of Sciences, Institute of  
Ethnology,  
Budapest, Hungary

e-mail: mseilyefalvi@gmail.com

### **Transdisciplinarity in digital scholarly editing? The perspective of a folklorist**

#### **Keywords:**

Digital Editing / Textualization / Folklore Texts /  
Transdisciplinarity

The textualization of folklore texts has been one of the most elementary issues since the beginning of the discipline. Broadly understood, folkloristic textualization involves more than merely taking down, transcribing, and annotating folklore texts. It also includes theoretical considerations before going to the field (what needs to be collected and how) and subsequently the selection and classification of texts, that is, the entire preparatory process for publication, are integral parts, including theoretical and methodological preconceptions. Folkloristics has undoubtedly arrived at a new paradigm with the possibility of digital textual scholarship. Although folklore studies have already produced innumerable attempts and a prolific literature on this topic, we are still at an early stage in terms of theories and methodology. Digital humanities declares itself ab ovo inter-, multi-, (and even) transdisciplinary, however in practice it is very hard – sometimes nearly impossible – to achieve it. How does this characteristic digital humanities change the way of scholarly editing? How can we keep in mind and cater to the different needs of different disciplines while making digital editions interoperable? For transdisciplinarity to truly work the needs of the various disciplines have to be considered ahead of time. My paper examines the challenges and possibilities the digitisation of folklore texts present in the wider context of digital humanities and the archiving and textualization strategies in folkloristics.

## Tamás Kiss

Central European University,  
Budapest, Hungary

e-mail: [tamas.mail@gmail.com](mailto:tamas.mail@gmail.com)

### **Introducing Ottoman Turkish Text Analysis Software Rumi 1.0: A Quantitative Analysis of Gelibolulu Mustafa 'Ali's Works**

#### **Keywords:**

Rumi 1.0 / Ottoman Turkish / Text Analysis / Text Mining / Gelibolulu Mustafa , Ali / Nusretname / Nushat Üs-Selatin / Künh Ül-Ahbar

Despite the vast corpus of Ottoman historical documents awaiting examination, computational text analysis and text mining practices are still almost never used in Ottoman Studies. Heretofore the main reason for this shortcoming of the Ottomanist field has been the lack of digital tools sufficiently applicable to texts written in the Osmanli language. Rumi 1.0, a software that I have developed to fill this gap, is the only software to date optimized to computationally analyze Ottoman documents. In this paper I will demonstrate the practical use of the software through analyzing the data extracted by Rumi 1.0 from sixteenth-century Ottoman historian, poet, and bureaucrat Gelibolulu Mustafa 'Ali's (1541-1600) oeuvre.

Exploiting the characteristics of the Osmanli language's unique lexicon (i.e. comprising of mainly Turkic, Persian and Arabic words) and morphology (i.e. one that represents a heavily inflected agglutinative language) through a combination of NLP and text mining techniques, Rumi 1.0 helps historians and literary critics identify the writing style and sociolinguistic as well as educational background of individual Ottoman authors. It is also a powerful tool to locate parts of documents borrowed from other authors or altered during the editing or copying process. At a larger scale of analysis, Rumi 1.0 may enable researchers to identify the writers of anonymous documents that thus far have remained unassigned to a specific author.

Mustafa 'Ali's works are an ideal testing ground for Rumi 1.0 for several reasons: 'Ali was one of the most fertile authors in the Osmanli language, whose volume of literary output provides an adequately large corpus fit for computational text analysis; As 'Ali's life trajectory is well documented, the results provided by Rumi 1.0 can also be measured against the author's biographical data; Besides his historical works such as Nadir ül-Maharib (The Rarity of Warriors, 1569) or the monumental Künh ül-Ahbar (The Essence of History, 1599), for the writing of which he had to resort to borrowing from earlier authors, 'Ali also wrote about his own era based on his first-hand experiences such as in Nusretname (Book of Victory, 1580) and Nushat üs-Selatin (Counsel for Sultans, 1581). Works of the latter kind allow for a more accurate overview of 'Ali's writing style, the quantitative measurements of which can be compared to sections of his composite works (e.g. chronicles) to identify which sections are likely to have been copied from other authors' works or subsequently interfered with.

In my paper I will account on each of these analytical steps, namely (1) a quantitative analysis of Mustafa 'Ali's writing style based on Nusretname and Nushat üs-Selatin, (2) measuring the data against subsections of his chronicles to identify passages likely to have been borrowed from another author, and (3) comparing the data with 'Ali's biographical information to test whether the results yielded by Rumi 1.0 provide historians with relevant information about an Ottoman author's linguistic and intellectual upbringing. Conclusively, I will (4) propose the tentative applicability of Rumi 1.0 in the mapping of Osmanli language usage in the early modern Ottoman Empire.

## Tibor Koltay

Eszterházy Károly University,  
Eger, Hungary

e-mail: [koltay.tibor@uni-eszterhazy.hu](mailto:koltay.tibor@uni-eszterhazy.hu)

**Digital humanities: opportunities and challenges for librarians, library and information science and information literacy**

### Keywords:

Librarianship / Library and Information Science  
/ Information Literacy

The proposed paper is about the challenges that librarians face when embracing digital humanities (DH) and argues that the underlying discipline of librarianship, library and information science (LIS) has the potential to provide theoretical background for these efforts.

The shared strengths of DH and LIS, both associated with the academic use of recorded information, substantiate optimism. For instance, the prevalence of data, defining important directions of curation and research data management is a motivating force for both.

Regarding disciplinary affiliation, the digital humanities are already using mathematical and statistical methods borrowed from computer science and the natural sciences. Therefore, they could also benefit from using methods of the social sciences, similarly to LIS, which is one of the social sciences, while many of its subfields address issues and use methods, taken from the humanities. This may bring in advantages in several undertakings in both disciplines, including the attempts to understand the effects of a growing digital infrastructure as a system for knowledge production, as well as acquiring a deep understanding of the mutual co-constitution of technology and of the human component.

There are several DH-related competencies, possessed by librarians, such as the ones related to the acquisition and disposition of resources, the maintenance of collections, including preservation and conservation and the usefulness of knowledge and skills in the provision of metadata by librarians is indisputable. Notwithstanding, there should be more space for sharing views about cooperation between digital humanist and librarians, as well, albeit librarians' contribution to DH projects can be described on a scale that begins with providing services that minimize the time that scholars must spend on technical and administrative processes. At the other end of the scale, there is the stage of becoming true collaborators. The need for achieving this, possibly with less friction between librarians and teaching staff members will be explained through examples.

The future development of DH is unimaginable without educating new generations of digital humanists. Such education requires equipping students with varied literacies, including information literacy and data literacy skills and the corresponding mentality. The broader framework of this, i.e. participation in DH pedagogy may become such an opportunity for libraries, while it faces several obstacles that also can be identified as varied types of overload, concerning tools, projects, debates, and pedagogy overload. These types of overload also will be addressed.

A related question is how can digital humanities instruction become more information-literate, the necessity of which has been mainly advocated by librarians and found little echo beyond the library world. This paper will argue for making use of state-of-the-art approaches to information literacy, especially those that draw more attention to the use of phenomenographic and sociocultural theories and discourse analytical approaches, because they are not only useful, but may be more readily accepted among DH scholars.

**Alexander  
König –  
Verena  
Lyding –  
Elisa  
Gorgaini**

Eurac Research,  
Bozen, Italy

e-mail:

[alexander.koenig@eurac.edu](mailto:alexander.koenig@eurac.edu)

[verena.lyding@eurac.edu](mailto:verena.lyding@eurac.edu)

[elisa.gorgaini@eurac.edu](mailto:elisa.gorgaini@eurac.edu)

**Building a digital infrastructure in South Tyrol**

**Keywords:**

[Infrastructure / Regional Development /](#)

[Language Resources / Cultural Heritage](#)

With this article we present the DI-ÖSS project, a local infrastructure initiative for South Tyrol, which aims at connecting institutions and organizations that are working with language data. The digital infrastructure serves to facilitate and foster data exchange, joint efforts in processing and exploiting data and the overall increase of synergies, and thus links to big European infrastructure initiatives. However, while sharing the overall objectives to foster standardization and increase efficiency and sustainability, on the implementation level a local initiative faces a different set of challenges. It aims to involve institutions, which are less familiar with the logic of infrastructure and have less experience and resources to deal with technical matters in a systematic way.

In this presentation, we describe how DI-ÖSS is addressing the need for digital language infrastructure on a local level; we lay out the course of action; and we depict the targeted mid and long-term outputs of the project.

In recent years, the field of Digital Humanities saw the development of multiple infrastructure projects at the European level. Among the most well-known initiatives CLARIN and DARIAH target the needs of researchers, with CLARIN being mostly centered

around the discipline of linguistics, and, to a lesser degree, history and literary studies, while DARIAH focuses on the broader field of all the arts and humanities.

Europeana, on the other hand, is focusing on the cultural heritage sector. Its main aim is to strengthen the networks between institutions like galleries, libraries, archives and museums (GLAM), especially by aggregating as much as possible their metadata to make them searchable in an easier and more convenient way.

The large sector of smaller research and public entities is not targeted by any of these big infrastructures, even though it could benefit from a close collaboration with Digital Humanities. This third sector contains smaller libraries, archives, cultural associations, and publishing houses, actors that deal with language and are peripherally contributing to the field of research and heritage, but who are themselves too small to easily participate in one of the big infrastructures. These minor but central players are the target of the DI-ÖSS project.

In contrast to the top-down approaches of the bigger projects like CLARIN or Europeana, DI-ÖSS is using a bottom-up strategy by following the actors that are producing and collecting linguistic data daily, which build the cultural and linguistic heritage of the region of South Tyrol. This entails getting into contact with potential language partners within South Tyrol to analyze the kind of language data they are working with; which and how much data they produce themselves; which are their typical workflows; and what could be potential synergies with other institutions.

Finally, we lay out the pilot phase, in which Eurac Research works closely with sample institutions to develop concrete use cases to showcase how such a local language infrastructure could enrich and optimize their work.

## Jessie Labov – Anton Mudrak

Central European University,  
Budapest, Hungary  
Blinken Open Society Archives,  
Budapest, Hungary

e-mail:  
labovj@ceu.edu  
MudrakA@ceu.edu

### **Encrypted Channels, Distributed Networks: The Telex footprint in the Cold War and its legacy for media practices today**

#### **Keywords:**

Networks / Text Analysis / Media History /  
Cold War / New Media / Science Technology  
Studies

The encrypted telex communication of Radio Free Europe/Free Europe Committee (RFE/FEC) represents an exciting new territory for research into Cold War history. Developed during the 1920s and 1930s as a supplement to the telegraph and the telephone, the telex – a distributed network which allowed encrypted communication between teleprinters -- was the primary medium used by governments, press offices, and large corporations for long-distance, secure communication from the 1950s through the 1980s. While there are several memoirs about RFE/RL by former practitioners (see Siefert 2003), and important recent scholarship by a new generation of historians (Kind-Kovacs 2015), there has been no significant work done on the telex communications, which were only recently digitally processed. This presentation will describe three levels on which our digital humanities project is operating: 1) new media archaeology and revisionist media ecology: following the work of STS scholars like Matthew Fuller (2005), Johan Söderburg (2006), and Maxigas Dunajcsik (2015) who trace the roots of contemporary media practices in older/parallel media from the early computing era. Here the telex can be studied as an early form of distributed network communication (anticipating email in its ‘cc-ing’ capabilities), which overlaps with early terminal-to-terminal computing exchanges and fax messaging. We are interested in the electronic and material practices that grew out of the telex system and where we might see their legacy is today. 2) Meanwhile, working with the metadata now available in the processed telexes (1960-1973, over 30,000 individual messages and 470 unique authors) we have begun an initial analysis of the distribution and frequency over time, by author, and mention of country. We will soon include entities mentioned in the telexes (already recorded in the meta-data). Using this matrix to establish a baseline of how the network functioned during periods of relative calm, we will select three ‘crisis points’ in the 13-year period already processed and analyze the changes we see in this network in response to pressures on the system. 3) The third level of analysis is contingent on the successful OCRing of the digitized images in order to render the text machine-readable. If this challenge is met, we will begin a series of text mining experiments, extracting directional values, running topic models, cluster analyses, and ultimately comparing draft versions of longer reports – sent and edited via telex – with RFE background reports and situation reports. The significance of this project lies in the disciplinary intersection of digital history, STS, and new media studies: we expect to show how the media ecology of the telex-verse shaped the communication and curation of information by RFE. By embedding this knowledge of the network in the chronological narrative of the Cold War, which has been exhaustively studied on a granular level, we should be able to account for directional changes as crises erupted from different sources. Lastly, the hybrid nature of information flow between media and across borders at a time when these spheres seemed relatively restricted could provide us a helpful model for understanding similarly mediated practices today.

# Davor Lauc – Darko Vitek

University of Zagreb,  
Zagreb, Croatia

e-mail:  
dlauc@ffzg.hr  
dvitek@hrstud.hr

## **From the History to the Story - Harvesting Non-Monotonic Logic and Deep Learning to Generate Multilingual Family Narratives From Genealogical Data**

### **Keywords:**

Digital History / Deep Learning & the  
Humanities / Natural Language Generation /  
Non-monotonic logics / Genealogical Record  
Linkage

Genealogical sources, like parish registers, represent an immense resource for family history and microhistories. However, it takes a significant amount of effort to accurately transcribe, clean, represent, and interpret data in a way that creates a narrative that can clearly and comprehensibly present data to wider, non-professional, audiences.

In this presentation, we will describe a methodology and computing system that has been developed to automate the extraction, cleaning, and transformation of the data contained within parish registers into a highly structured semantic web format. Genealogical data that is represented in this way can be used for both historiographical analysis and knowledge dissemination using natural language generation techniques. The system was developed and tested using 19th century Roman Catholic Church baptism and death records for the island of Vis, Croatia. As automated handwriting recognition technologies are still not up to this task, the transliteration was performed manually, resulting in records covering 26350 events and 86422 personas (mentioning of the persons). Using a deep learning sequence to sequence the models, the annotated textual data were cleaned and converted into a semantic web format. In the interest of data consistency and interpretability, object values were used instead of text values whenever possible. This implies, for example, that the parts of the individuals' names, like first names, surnames, nicknames, and so on, represent first-class citizens within our ontology. The person records only refer to these entities, enabling consistent and centralized translations and quality management. Semantic web claims are stored in the Wikibase environment, which facilitates crowdsourced data enrichment from diverse additional sources. The central, most challenging subsystem is a reasoning engine that automates the process by which inferences are drawn from the additional information that is implicit in genealogical data. The main part of this subsystem is a link prediction system that infers two classes of relationship among personas: equivalence, the task known as record linkage, and a kinship relation. Record linkage is performed by two interconnected algorithms—the entity retrieval and similarity algorithm, and non-monotonic inference algorithm. The task of the former is to retrieve the most similar personas. It is based on a dense vector representation of the entities that is learned by Siamese neural networks on a set of labelled pairs. As every possible decision on persona equivalence implies other equivalence and kinship relations, the task of the latter algorithm is to infer all consequences of such choices and include them in the loss function of the neural network. This system is based on the non-monotonic answer-set system. Once the data has been cleaned, highly structured, and implicit information is inferred, it is possible to create narratives using techniques of natural language generation. In our showcase for the Vis Island, three books that contained over 800 pages each were generated in the English, Italian, and Croatian languages.

**Daniel  
McDonald –  
Eveline  
Wandl-Vogt –  
Mahsa  
Vafaie**

Bitpanda GmbH,  
Vienna, Austria  
Austrian Academy of Sciences,  
Vienna, Austria

e-mail:

daniel.mcdonald@oeaw.ac.at  
eveline.wandl-vogt@oeaw.ac.at  
mahsavafayi@gmail.com

**Integrating crowdsourcing and the blockchain  
for natural language data collection: novel  
methods, use-cases and debates**

**Keywords:**

Blockchain / Crowdsourcing / Open Science /  
Open Innovation / Digital Literacy / Open  
Access / Software / Methodology / Data  
Collection

Already, vast amounts of digital data, much of it linguistic and culture in nature, is collected via computer-mediated communication and crowdsourcing systems. Unresolved issues inherent to such approaches include access and ownership of produced data, potential for homogenisation of human language and culture, development of new forms of exploitation and control, and a general lack of consideration of the centrality of interpersonal and ideational context the meaning-making process. Often, inappropriate task design, assignment provision, and/or reward size have led to low-quality responses, of little use for research or re-use. One nascent technology that may enhance the utility of such approaches is the blockchain [1]—a decentralised, trustless ledger that can accurately keep track of digital information. To date, the most common use-case for blockchain technology is as a currency or payment network (e.g. Bitcoin, Ethereum). Blockchain-based cryptocurrencies permit the transfer of real or symbolic value in a way that is very resilient to system outages and malicious code. Meanwhile, blockchain-based databases are provably open-source, limiting researcher bias, increasing reproducibility, and promoting data re-use [2]. In this paper, we present a prototype for blockchain-based, open-access, contextually sensitive collection of real language and cultural data, pioneered within a current open innovation exploration space at the Austrian Academy of Sciences [3,4], aimed at facilitating open DH research. Language questions are generated from a preexisting database and turned into short visual games submitted to the most suitable user agents; returned answers are added to a decentralised database hosted by those who wish to use the data for downstream tasks. A consensus algorithm, coupled with the completeness of user profiles and prior history of response quality, determines the size of a reward, released to each user's account in the form of an Ethereum-based cryptocurrency [5]. Finally, an open-source API allows querying of the generated data, and dynamically presenting interesting insights online in real-time. The initial datasets are not perfectly clean; as such, the system is also designed to aid in the manual correction of poorly digitised historical documents. Because games put to the crowd are tasks on which machines score poorly (e.g. identifying commands in declarative/interrogative clauses, or deciding if a proper noun refers to a person or place), and because the system maps answers to anonymous user profiles, the collected data has potential use across AI, NLP, social science and DH research alike. Depending on the initial dataset used, the system could be equally well-suited to collect data suitable for use in tasks such as linguistic typology (in classifying languages and dialects), computational linguistics (i.e. in natural language generation and parsing), the social, political and population sciences (in mapping language use to demographic details, or uncovering attitudes toward the data shown to participants). Qualitative engagement with veteran system users is also planned, in order to build a more complete picture of the lived experiences of user-agents, and their motivations for participating.

## Berrie van der Molen

Utrecht University,  
Utrecht, Netherlands

e-mail: [berrie.j@gmail.com](mailto:berrie.j@gmail.com)

**The public framing of MDMA in Dutch cross-media debates. Developing a digital tool to answer historical research questions based on distant and close reading of cultural heritage “big data”**

### Keywords:

[Big Data](#) / [Public Debate Research](#) / [Distant Reading](#) / [Mdma](#) / [Tool Development](#)

Since the digital turn (Nicholson 2013), it has been possible to do “bottom-up” keyword search in large amounts of digitized newspaper data. In CLARIAH research pilot DReAM (Debate Research Across Media) we have contributed to the development of a tool that simultaneously searches the digital Royal Library’s newspaper dataset and the digitized television and radio archive of the Netherlands Institute for Sound and Vision. We worked to enable public debate research across both textual and audiovisual cultural heritage documents in a fusion of digital methodology with a historical research question. This was a challenge, particularly because data-driven media research tools tend to affect research questions (Bron et al 2015).

My historical research interest is in public debates on drugs and the historical interaction between these debates and drug regulation between 1945 and 1990. One substance of interest is MDMA, perhaps better known as party drug ecstasy. It was introduced into public debates in the mid 1980s in The Netherlands. Despite its origins in medicine, the substance failed to acquire a significant medical role at any point.

Recent renewed attention for the role of MDMA in the treatment of Post Traumatic Stress Disorder (e.g. Amoroso and Workman 2016) may start to push the substance beyond its current recreative (non-medical) reputation. In order to understand MDMA’s initial public framing in The Netherlands, I study historical Dutch media debates in newspapers and on radio and television.

In this study I combine distant and close reading techniques in what we have called the leveled approach (Van der Molen et al 2017). The benefits of distant reading help to unlock the historical potential of huge datasets, while the relevant material is still subject to historical interpretation (“close reading”). The aim of research pilot DReAM was to accommodate this strategy for both print and audiovisual datasets. The potential benefits are numerous: it creates opportunities for enriched, fine-grained analysis on a large scale over long periods of time across media; a lot of time could be saved by not having to manually search for all the relevant material; material that may have stayed buried in huge datasets could rise to the surface.

This presentation will focus on the opportunities and challenges of extending the leveled approach to audiovisual material. How can I ensure that using distant reading techniques yields results that can produce a credible (cultural history) narrative? Although the potential of semantic text mining for historians could be huge, there are still hurdles to overcome here (Snelders et al 2017). And how did we choose to deal with the datasets’ different (and uneven) access and Intellectual Property Right contexts? While presenting the proceedings of research pilot DReAM, the talk will keep an eye on the historical research question regarding the public framing of MDMA.



**Majlinda  
Muka –  
Dorina  
Xheraj-Subashi**

Aleksander Moisiu University,  
Durrës, Albania  
Center of Albanological Studies,  
Tirana, Albania

e-mail:

elamuka@hotmail.com

dorina.xheraj@hotmail.com

**The influence of digital storytelling on the  
achievement competences of students in  
cultural heritage teaching**

**Keywords:**

Digital Storytelling / Cultural Heritage /  
Achievement Competencies

Digital storytelling is a relatively new educational approach in the contemporary literature of teaching methodology which results in positive performance indicators of student achievement. The purpose of this research study is to investigate and evaluate the effectiveness of applying the digital storytelling approach to heritage culture curricula teaching in a middle school. We state the hypotheses that the application of digital storytelling during the implementation of teaching in cultural heritage course positively influences the knowledge achievement obtained by students. A group of 68 students, grade 6 in a middle school in Durrës, Albania, underwent in observation for a one month period during the performance of the classroom management teaching activities. The group was divided among 33 students as control group managing based conventional teaching activities and a 35 students group in which digital storytelling technique was applied to teaching activities. At the end of the experimentation period, the groups underwent an achievement test. The results showed that the implementation of digital storytelling approach in the cultural heritage course teaching activities had statistically significant impacts on the performance of students' achievements among groups.

## Anna Neovesky – Frederic von Vlahovits

Academy of Sciences and Literature Mainz,  
Mainz, Germany

e-mail:

[anna.neovesky@adwmainz.de](mailto:anna.neovesky@adwmainz.de)

[frederic.vonvlahovits@adwmainz.de](mailto:frederic.vonvlahovits@adwmainz.de)

**IncipitSearch: A common interface for  
searching in music repositories**

**Keywords:**

[Digital Musicology / Metasearch / Data  
Aggregation / Open Data](#)

Open research data is facilitating broader ways of using, reusing, enriching, and linking research results. Many services use metadata to connect information from different platforms and repositories. We created a metasearch that builds up on annotated music to connect musicological collections and repositories.

As early as the 1960s, Music librarians introduced the idea to generate a human and machine readable standardised format to identify music by its melodic beginning, the incipit. Barrey S. Brook and Murray Gold developed the Plaine & Easie Code that allows the transcription of the beginning notes of a musical piece into a combination of numbers and letters. (Brook and Gold, 1964)

Music information retrieval systems build up either on audio or symbolic music notation. In digital musicology, that deals with notation and critical digital edition of works, the search in notated music is an important application case (Typke et. al. 2005).

RISM is undoubtedly the most established repository for musical data. It contains over one million records of historic music materials and over 1,7 million musical incipits (for manuscripts only), which can be accessed using an incipit search ([http://www.rism.info/en/home/newsdetails/select/rism\\_online\\_catalog/article/2/music-incipit-searches.html](http://www.rism.info/en/home/newsdetails/select/rism_online_catalog/article/2/music-incipit-searches.html)). But other incipit repositories exist that cannot be accessed via RISM because they either have not been implemented as data yet or because they are not a type of resource the RISM collection is focusing on and will not be added to the catalogue, such as work catalogues.

IncipitSearch is a tool and a service specifically tailored for research on music incipits. It is simultaneously a centralised data endpoint where multiple aggregated catalogues of incipits can be accessed as well as a decentralised open source software that can be integrated as stand-alone search in other platforms. A microservice based software architecture allows high flexibility in usage and extension of individual components (Haft, Neovesky and Reimers 2016).

IncipitSearch enables users to enter search queries in the search field by playing them on a virtual piano keyboard while Plaine & Easie Code can also be directly entered into the search field. Search with transposition or with exact matching can be selected (<https://incipitsearch.adwmainz.net>). Next to the found concordant incipits, the result list displays backlinks to the entry in the respective catalogue. Through consistent usage of authority control and metadata standards, IncipitSearch is an open source tool and service warranting sustainability, transparency, and accessibility of research data.

**Thomas  
Palfinger –  
Alexander  
Schatek –  
Amelie  
Dorn –  
Eveline  
Wandl-Vogt**

Austrian Academy of Sciences,  
Vienna, Austria  
Topothek,  
Vienna, Austria

**Food cultures: co-creation and evaluation of a  
thesaurus as a cultural infrastructure**

**Keywords:**

Academia-Society Collaboration / Co-Design /  
Co-Creation / Cultural Heritage / Food History /  
Food Cultures

Close collaboration with societal actors is a central part of the Digital Humanities project exploreAT!(exploring austria’s culture through the language glass). Digital tools have increasingly enabled the process of interconnecting and opening up various resources across different media. To make use of this potential, exploreAT! cooperates with the Topothek, a citizen driven digital infrastructure which collects local historical multimedia items from private sources. It is of interest to connect this collection to exploreAT! as it contains a variety of cultural and local historical information, which can only be accessed with the help of citizens and their knowledge. To initiate such collaboration, both scientists and citizens need to learn how to interact with each other in order to constitute common goals. In this paper, we focus on the workflow and process established. We introduce the community group “Topothek”, a citizen driven archival platform to archive, annotate, geo-reference and present regional culture. We present our first co-created result, a food cultures thesaurus, interlinking food terms with cultural practices and tools. The added value for Topothek is increased discoverability, while for exploreAT! it is the scientific exploitation for cultural lexicography of the multimedia archive. The processes along the mutual learning scenarios are evaluated. The co-created thesaurus is co-funded by the DARIAH-theme 2017 “public humanities” in collaboration with ADAPT centre Dublin (Ireland).

**Ioulia  
Pentazou -  
Ioanna  
Laliotou**

University of Thessaly,  
Volos, Greece

e-mail:

pentazou@yahoo.com

laliotou@uth.gr

**Digitizing the History of Academia: Creating  
the Digital Archive of the University of  
Thessaly**

**Keywords:**

Digitization / Digital Archive / Documentation /  
Cultural Heritage

Digital era has led to a renewal of interest in archive and archiving practices as new sets of relevances come into prominence, which can lead to the generation of new connections and a different set of classifications. As archive stands behind all research, we need to enquire into the processes leading to the reconceptualization of the archive in both theoretical and practical aspects. Questions of organization, storage and access require new methods of handling information and documents in the light of mass processes of digitalization. Within this context, University of Thessaly launches a new project of developing a central digital archive aiming both in the preservation and organization of different kind of records and documents. The project aims at creating two new innovative integrated web-based platforms: one for preservation-management-dissemination of archival digital documents and one for text and data mining of these documents. The archival items are connected with the history and operation of the University of Thessaly and contribute to the research, study and promotion of the Thessaly region cultural heritage. Main objectives of the project include the development and the full function of the above-mentioned platforms as research tools, the creation on new digital cultural heritage archival documents, the development of new digitisation and cultural heritage documentation technologies, techniques and methods concerning materials from the University of Thessaly and the region's history.

## Róbert Péter

University of Szeged, Department of English Studies,  
Szeged, Hungary

e-mail: [robert.peter@ieas-szeged.hu](mailto:robert.peter@ieas-szeged.hu)

### **Metadata visualization of 18th-century British newspaper and periodical articles concerning Freemasonry**

#### **Keywords:**

[Bibliographic Metadata / Visualization / Quantitative Analysis Of Newspapers / 18th-Century Studies / Freemasonry](#)

Despite its alleged secrecy and the performance of rituals in private space, eighteenth-century freemasonry was more visible and transparent in the public sphere than the present-day fraternity. Newspapers of the period not only published reports about grandiose masonic feasts, processions, concerts, charitable activities, laying of foundation stones and forthcoming lodge meetings but also informed their readers about the names of the newly initiated and foreign visitors of lodges. Often these articles provide the sole information about a particular lodge, event or member because of the loss or lack of original records including lodge minutes or membership lists. The objective of this paper is to highlight the ways in which the systematic analysis of the hitherto largely unmapped masonic-related press articles with digital methods can facilitate research on the history of British freemasonry from a wide range of perspectives, nationally and internationally. For example, it helps researchers to explore the changing public and self-perceptions of British Freemasonry as well as its influence on British associational culture during the long eighteenth century, and contributes to the understanding of the role of freemasonry in British social, cultural and political history. The analysis draws on a database of Irish, Scottish and English masonic press material, which contains the bibliographic metadata of almost 12000 articles between 1709 and 1813. A selection of this press material was published in the fifth volume of *British Freemasonry, 1717-1813* (Routledge, 2016). The paper briefly introduces the AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) web service in the making, which was used for the analysis. The AVOBMAT offers advanced search functions as well as basic and complex visualizations of datasets and library repositories. This new digital tool helps researchers to explore long-scale trends and patterns in historical, literary, linguistic and cultural processes. Its functions include the interactive visualization of the author-publisher(-bookseller) networks, the identification of female and male authors (the gender is inferred automatically from the first names) of the uploaded database and the visualization of their distribution over time.

**Opposite-Sex Relations in Present-Day Dutch Literature. A Network Analytical Approach to Character Representations**

**Keywords:**  
Social Network Analysis / Opposite-Sex Relations / Digital Literary Studies / Dutch Literature

Relations between characters with certain demographic profiles have lain at the basis of ideological approaches to literature (e.g. Pattynama 1994, Murfin 1996, Meijer 1996, Pattynama 1998, Minnaard 2010, Sidi-Said 2010, Song 2015, Mushtaq 2010, Fatima et al 2015.). The literary representation of gender, race or class is frequently analysed through a critical examination of the ways in which e.g. male and female characters are represented in interaction with each other. Through critical close readings, hierarchies can be discerned that pertain to the respective dominant or peripheral position a certain character takes up in a literary text.

Notions as ‘dominance’, ‘importance’, ‘centrality’ and ‘conflict’ can be operationalized for literary analysis through the method of social network analysis. A recent development in data-driven research on literature and culture is the analysis of fictional networks (Alberich e.a. 2002, Stiller e.a. 2003, Elson e.a. 2010, Lee & Yeung 2012, Karsdorp e.a. 2012, Agarwal e.a. 2013, Jayannavar e.a. 2015, Karsdorp e.a. 2015, Lee & Wong 2016). Through the analysis and interpretation of character networks, the present research aims to explore character representations in present-day Dutch novels in an empirically informed way. In order to do so, we combine a network analytical approach with an ideologically oriented close reading.

We focus on a sample corpus of 170 recent Dutch novels containing  $\pm$  1600 characters, of which extensive demographic information (gender, age, place of birth, place of residency, level of education, profession) has been gathered (Van der Deijl, Pieterse, Prinse, Smeets 2016), as well as thematic roles (family, lover, colleague, friend, enemy) between all characters. Nodes in the networks are represented by the characters in the novels, edges by the relations/interactions between those characters. The weight of the edges is measured through co-occurrence frequencies in shifting windows of N-words. For each narrative mode (first person narrated, third person narrated, narrated by multiple narrators) a customised method to measure the weight is used.

In this talk, we will present and evaluate the results of a case study of opposite-sex relations in the corpus. Relational patterns between characters of opposite genders will be discussed in light of close readings of specific novels that conform to or deviate from those patterns. We formulate the hypothesis that the relations between male and female characters comply to certain stereotypes (e.g. male characters are mostly related to female characters as lover or family, as opposed to colleague, friend or enemy). In order to test that hypothesis, we focus on the thematic roles through which opposite-sex relations are represented. Furthermore, we intersect the gender categories with demographic categories as age, descent and education. As such, we will give an insight into relation types and the demographic categories that play a significant role in the representation of opposite-sex relations in present-day Dutch literary fiction. We evaluate our results by zooming in on the representation in specific novels of a.o. young female versus old male characters, non-Dutch female characters versus Dutch male characters, lowly educated female characters versus highly educated male character.

## Filomena Sousa

Memória Imaterial,  
Portugal

e-mail: [filom\\_sousa@yahoo.com](mailto:filom_sousa@yahoo.com)

### What about Digital Platforms of Intangible Cultural Heritage?

#### Keywords:

Digital Humanities / Intangible Cultural Heritage / Inventory / Digital Platforms / Communities

One of the main objectives of inventorying elements of the Intangible Cultural Heritage (ICH) is the systematized register and organization of knowledge produced in respect of this heritage. Article 12 of the Convention (UNESCO, 2003) refers to the need to promote the ICH inventory and all elements applying for enrolment in one of the two worldwide lists of ICH have, necessarily, to be listed in an Inventory. Mostly for this reason, in the last 10 years multiple processes of inventory have begun and, in order to comply with the recommendation of public access, resorting the Web became a logical option. In this paper is presented the The Map of e-Inventories of Intangible Cultural Heritage that enables direct access to 158 e-inventories through the links marked in countries where these were carried out. This paper presents a preliminary analysis of some of the data that characterize these inventories.

The construction of the Map of e-Inventories of Intangible Cultural Heritage aims to identify and collect data of Digital Platforms available on the Web and dedicated to this type of heritage. The definition of “ICH e-inventory” adopted by the project includes: online published inventories with free access, of or about expressions of one or more domains of the intangible cultural heritage – oral traditions and expressions (including language as a vehicle of the ICH); performing arts; traditional craftsmanship; social practices, rituals and festive events; knowledge and practices concerning nature and the universe (UNESCO, 2003). For this purpose, are not considered: inventories in preparatory stage without available content; web sites with exclusively touristic information or denominative lists in which do not appear, at least, a description or transcription of the cultural expressions included in the inventory.

The Map enables direct access to the e-inventories through the links marked on their respective countries and for each inventory, it displays the specified related information: format (database, website, or PDF); geographical level; promoter entity; domains of the ICH; main fields of the inventory; multimedia resources; language and associated social networks.

With the publication of the prototype of this project, it is intended that the Map of e-Inventories of ICH should be subjected to reviews of collaborative methodology. As a work in permanent construction, authored and coordinated by Memória Imaterial – a Non-Governmental Organization (NGO) accredited by the UNESCO and based in Portugal –, it must be enriched by any additional information to update the mapped data. With the aspiration to promote this collaborative network we invite all practitioners of cultural expressions, heritage professionals (from public and private institutions), local associations, researchers, NGOs, and other interested parties on issues of cultural heritage, to share information on inventories of the ICH.

## Botond Szemes

Eötvös Loránd University,  
Budapest, Hungary

e-mail: boboszemes@gmail.com

### **The practice of the close reading, and its possible alternatives**

#### **Keywords:**

Close Reading / Digital Age / Combinatoric Reading / Distant Reading / History Of Science

My presentation is a kind of a history of science, especially the history of the literary studies. The question in the center would be „what does scientific reading mean, and how does it change in the age of digitalization?” Therefore I would like to show the historical, medial, social, educational and aesthetical context of the concept and the practice of close reading – through re-reading the most important texts of the New Criticism. After that I would like to make an order in the chaotic mass of the concepts from the last two decades. These articles (mainly from the anglo-american world) detect, that new, alternative ways of reading have emerged, and/or could be fruitfully used in the scientific practice. The biggest aim of the presentation is to see the contexts behind these ideas, and to show the similarities and the differences between them. It can be interesting and informative to see, how far the distant reading from the close reading is, for example. In this sense it would be a comparative approach. I think, this is a necessary step to establish new kind of practices, and maybe to renew the literary criticism. I believe that these alternative ways are not excluding each other, but can be connected for the sake of the actual research.



**Miklós  
Tamási –  
András  
Török –  
Bettina  
Fabos**

Fortepan,  
Budapest, Hungary  
Summa Artium,  
Budapest, Hungary  
University of Northern Iowa,  
Cedar Falls, Iowa, United States

e-mail:

fortepan@gmail.com  
andras.torok@summa-artium.hu  
fabos@uni.edu

**Duplicating the Fortepan Photo Archive in  
Neighboring Countries**

**Keywords:**

Photo / Archive / Hungary / Creative Commons

Fortepan (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). 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 presentation will illustrate nearly a decade of Fortepan (successes as well as failures), sharing what we have learned, and discussing further collaboration with other countries.

**Bogata  
Timar –  
Erika  
Asztalos –  
Ditta  
Szabó –  
Nikolett  
F. Gulyas**

Eötvös Loránd University,  
Budapest, Hungary

e-mail:

timar.bogata@gmail.com  
aszterik@gmail.com  
szadibotta@gmail.com  
kecskebuka@gmail.com

**The Typological Database of the Ugric  
Languages: establishing a long-term project**

**Keywords:**

Linguistic Typology / Database / Uralic  
Languages

The research project “Typological database of the Ugric languages” was accepted for a three year funding period by the Hungarian Scientific Research Fund in July 2012, and was successfully closed in September 2015. The extension of the project, „Typological database of the Volga Area Finno-Ugric Languages” was accepted in July, 2017, for a period of four years.

Our database (known in short form as the Ugric Typological Database, UTDB) is intended to display the typologically significant features of four languages/dialects of the Uralic Language Family. Each of these belongs to the Ugric sub-branch within the Finno-Ugric branch of Uralic. The varieties are the following: Hungarian (standard variety), Mansi (Northern dialect), and Khanty (represented by its two rather distinct dialects: Synja and Surgut). By 2021, the range of languages will have been expanded with the addition of Udmurt, Meadow Mari and Komi-Permyak. The typological database serves as the intersection of general linguistic and Uralic-specific (e.g. an intragenetic) typology, and is accessible to both researchers in the field of Finno-Ugric studies and linguistic typology. It is the first to provide typological analysis of the Ugric languages in terms of morphonology, morphology, morphosyntax and syntactic-level features, which are both structural and word-order related. For the purpose of this database, these features are called parameters, of which it currently features ca. 210. The ultimate purpose is to create a complete database of the Uralic languages, thus facilitate the exploration of relationships between these languages and the rest of the world’s languages in terms of their typological features.

The structure of the UTDB was inspired by the World Atlas of Language Structures (WALS) (Dryer & Haspelmath 2013), with adjustments to our needs and budget. Users are able to simultaneously search the database for multiple dialects, parameters, or parameter values. Our presentation will focus on the buildup and usage of the database, highlighting its assets and purpose. Since it is scheduled to almost to double in size by three additional languages, we will also mention the challenges of expansion and current maintenance.

Utdb.elte.hu has been available for all audiences since September 2015 (Havas et al. 2015). It is currently accessible in Hungarian, English and Russian to full extent. Minor corrections can be undertaken at any time.

**Yun-Cheng  
Tsai -  
Pu  
Yu -  
Chia-Ching  
Wu -  
Ji-Yu  
Huang**

National Taiwan University,  
Taipei City, Taiwan

e-mail:

pecutsai@ntu.edu.tw  
tobyupu@gmail.com  
jeremywu871205@gmail.com  
p98733211@gmail.com

### **Natural Language Processing Real-Time System for Central European News**

#### **Keywords:**

Named Entity Recognition (NER) / Natural  
Language Processing (NLP) / Real-Time System  
/ Sentiment Analysis / Text Mining / Data  
Visualization / NLTK Tree / Information Retrieval

Internet as a medium for information transmission by the media has become one of the main sources of people's knowledge absorption. People always get a preliminary understanding of a topic from the online media. In Central Europe, an incident or a certain issue affects more than one country, possibly multiple countries. Due to the different political and national conditions in various countries, the coverage of the terms and expressions described in the articles of various countries and regions in the news media will present a description and reflection of a certain issue. However, current online information become volume, velocity and variety to require specific technology and analytical methods for its transformation into value. In order to expand to other complementary characteristics of big data in social sciences and humanities (SSH) area.

We propose a natural language processing real-time system to analyze text information from online news websites automatically. The system uses the existing text analytics resources that are defined for collections. Our purpose is to perform analytics on any text documents. Users can immediately check the analysis visualization results. Our system has the following main functions for SSH researchers.

1. Collect information from online news websites or offline corpus from various countries and analyze the opinions on specific topics.
2. Researchers can use Named Entity Recognition (NER) classification of groups manually to define the information retrieval features, and use text patterns for sentence interpretation emotional markers.
3. The terms and phrases described in their articles will present descriptions and reflections on a topic.
4. Provide a platform that can instantly find and analyze a large number of texts.

The system will help SSH researchers have over view from their documents. Then they can get important information from text quicker than without using any technical supporting.

## Lucia Vannini

Institute of Classical Studies,  
London, United Kingdom

e-mail: [lucyvannini@gmail.com](mailto:lucyvannini@gmail.com)

### **Crowdsourcing projects in Classics: a reflection on models of collaborative editing of primary sources**

#### **Keywords:**

Digital Scholarly Editing / Collaborative Editing / Crowdsourcing / Teaching / Digital Classics / Citizen Science / Community Sourcing

Upon reflecting on the influence the new digital methodologies are having on everyday research practices and scholarly communication, academic crowdsourcing, i.e., the use of public participation to contribute to projects, is a method that can be considered as worthy of a specific analysis, as it is changing the way academic communities collaborate and communicate.

Rather than presenting the application of crowdsourcing in a specific project and its results (an aspect that academic literature on this method is already increasingly dealing with), I will seek to identify common trends among crowdsourced resources, from the point of view of the typologies of users involved. My analysis will focus on resources for Classics, a field for which an overall examination of collaborative resources has not been carried out yet, and will try to answer the question of how these place themselves in the broader context of crowdsourcing in the Humanities – in other words, what distinctive features crowdsourced projects for classical scholarship present, and what contribution can they offer to projects of this kind in other research domains.

I will therefore provide an overview of Classics crowdsourcing, and pinpoint the models of collaborative editing implied. I will start from a unique case in Classics, represented by the Ancient Lives project: this was explicitly designed for non-experts and devised to involve any interested individual, even with no training in Papyrology and Greek, in the transcription of papyrus fragments. I will then move on to exemplify a different model of crowdsourcing, known as “community sourcing”, which, though presenting a membership potentially open to everyone, de facto addresses to those who have the necessary expertise to edit or annotate ancient texts. Finally, the presentation will discuss some collaborative editing platforms that can be considered as intermediate between the two previous ones, in that they ask for contributions from groups of university students; their use is in fact incorporated into the teaching of ancient languages and literature and practiced in a classroom setting. For example, students’ contributions have been used in classes of Greek language and literature, as well as Medieval Latin, to build up a linguistic corpus (the Ancient Greek Dependency Treebank) or collections of manuscripts (the Homer Multitext and the Tufts University Library catalogue).

The aim of this analysis is to show that crowdsourced collections in Classics present two peculiarities. Firstly, they mainly address specific groups of participants, i.e., communities of scholars or students, rather than the general public through an open call recruitment policy (as much as with the significant exception of the Ancient Lives project). The embedding of crowdsourcing into the teaching of language and literature, which implies the recognition of a pedagogical value in this method, is the second specificity. I will discuss how the latter practice arises pedagogical questions on the evolution of teaching methods and on the relationship between teaching and the publication of scholarship, as students, including undergraduates, are enabled to actively participate in the creation of scholarly resources, thereby beginning to take part in professional research.

## Fruzsina S. Vargha

Eötvös Loránd University,  
Budapest, Hungary

e-mail: fruzsa@gmail.com

### **Estimated acoustic parameters from digitised dialect atlases**

#### **Keywords:**

[Hungarian Dialectology / Dialect  
Geography / Digitised Dialect Atlas / Vowel  
Formants / Acoustic Map](#)

In sociolinguistics, dialect geography and related fields F1 (first formant) and F2 (second formant) are used to determine phonetic vowel quality. However, such acoustic studies are rare in Hungarian dialectology (Vargha 2013, Kocsis 2013, Presinszky 2016), thus we do not have an overall picture about the geographic differentiation of the acoustic features of Hungarian vowels. The question arises whether vowel qualities comparable to those obtained from acoustic measurements can be deduced from the large quantities of impressionistic phonetic transcriptions gathered by scholars during the last 70 years and contained in printed dialect atlases. This paper will present a digital solution to this question, based on a quantitative method. In recent years more than one million data instances from several Hungarian dialect atlases have already been computerised in a series of interrelated computational dialectology projects. Unlike the original paper-based data, the newly created digital data sets can be searched, analysed and mapped automatically with appropriately developed linguistic software tools. In recently processed oral corpora, time-aligned phonetic transcriptions connect the original sound material to its transcribed form. Approximately 460 hours of linguistic interviews conducted more than 50 years ago for The Atlas of Hungarian Dialects at 352 locations have been digitised in 2005. Part of this material, extracts from spontaneous speech from 77 locations (published in the Hungarian Dialectal Talking Book series) have been already transcribed with time aligned phonetic transcriptions (Vargha 2007, 2011). These transcriptions were made according to the Atlas' standards thus can be used to determine, through formant measurements at the corresponding portions of the sound material, reference F1 and F2 values for each phonetic vowel type, usually transcribed as a combination of a vowel symbol and optional diacritical marks. The present paper is based on the quantitative analysis of the computerised version of The Atlas of Hungarian Dialects (1162 maps, 395 locations, 565330 data instances) and the acoustic investigation of the Atlas' digital sound recordings. Typical F1 and F2 values are calculated for each vowel type from its different phonetic representations (the different complex symbols referring to the same vowel), taking into consideration the number of occurrences at each investigation point. The results are presented on colour maps using progressive shading from the highest to the smallest F1 and F2 values attributed to a given sound. The outcome will be compared to the results of other quantitative analyses on the same corpus such as MDS dialectometric maps showing the dialect continua.

## Melinda Vasari

Eötvös Loránd University,  
Centre for Digital Humanities,  
Budapest, Hungary

e-mail: [vasari.melinda@gmail.com](mailto:vasari.melinda@gmail.com)

### **Securing the Literary Evidence. Some Perspectives on Digital Forensics**

#### **Keywords:**

Digital Media Environment / Digital Writing /  
Nachlassbewusstsein / Estate Consciousness

The focus of my broader research is the impact of the new medial environment of today on the practice of literary studies. The reason of my choice lies in the fact that digital technology has become unavoidable and indispensable both in our everyday activities and professional practices. My main question is how literary studies, especially philology can adjust to, apply, and utilize this new, digital media environment. The motivation of my research is to reinterpret, in the light of this, the already existing practices, by which I mean both the application in practice and the theoretical reflection. From a broader perspective, in short: as Kittler investigated how our concept of writing had changed by the appearance of the typewriter, and what kind of consequences it had on our practice of writing in everyday life and in literature, I examine these questions in regard to digital writing.

In this presentation, through three examples, I introduce how digital technology has changed the process of writing, and the practice of archiving itself that manifests in the different degrees of the so called “estate consciousness” (*Nachlassbewusstsein*). These three techniques of archiving technology that I introduce, require different philological practices that demand proficiency in digital technology on different levels. The most extreme case is when the philologist reconstructs the digital manuscript by data recovery, thus by applying the techniques of digital forensics for philological purposes. Since the process of writing (corrections, rewritings, extractions, temporal layers etc.) can be reconstructed by software, and this way, the genetic presentation of a given text is possible. This digital manuscript can be analysed by applying previously existing philological methods and, at the same time, by developing a special methodology that considers the new medial environment. Furthermore, it is important to emphasize that the philological work is preceded by the preservation, curation and processing of digital heritage and making it available for research – namely, the curation of literary archives that has been also transformed by the appearance and then by the dominance of digital media. I introduce the challenges that emerge as a consequence through the work of the Deutsches Literaturarchiv Marbach that curates Kittler’s digital heritage. All of this has theoretical relations, as well, since it influences how we understand the practice, the task and the scope of literary studies – these questions will be outlined by examples during my presentation.

## Cristina Vertan

University of Hamburg,  
Hamburg, Germany

e-mail: [cristina.vertan@uni-hamburg.de](mailto:cristina.vertan@uni-hamburg.de)

### **Multilevel Annotation of Historical Documents**

#### **Keywords:**

[Data Model](#) / [Annotation](#) / [Historical Corpus](#)

In this paper we will introduce a novel framework for data modelling which allows implementation of tailored annotation tools for the specific DH-project. We will illustrate the generic framework model by mean of three examples from completely different domains each treating another language: the construction of a diachronic corpus for classical Ethiopic texts; the annotation of classical Maya database of inscriptions and texts and the computer –based analysis of original and translation in three languages of historical documents from the 18th century. We will present the generic model and show the derived data model for each of the 3 examples and we will discuss the challenges implied by the development of a new software. We will illustrate also how interchangeability with other digital resourced is secured.

## Sara Wagner

Eötvös Loránd University,  
Budapest, Hungary

e-mail: saraberta.wagner@gmail.com

**Cultural heritage, digital access: the DIY archiving process of the Sirius band (1969-1973) and what it reveals**

### Keywords:

Digital Databases / Semi-Underground Status / Do-it-Yourself Preservation / Cultural Practices

Since the primary sources of the Hungarian popular music studies were created during the socialist era, both the creation and the preservation of the documents has been shaped by the socialist cultural policy.

The focus of my presentation is the digitally available Hungarian and Australian documents, mentioning Sirius, a progressive rock, jazz-rock band, which played in the socialist era. They were active roughly between 1969 and 1973 and spent a year in Australia. The band's fragmented presence in the popular music field, both before and after 1990, can be connected in curious ways to cultural policies in curious ways both in socialist and in capitalist Hungary.

Documents are from professional public archives such as Arcanum and Hungaricana, also the database of the National Australian Library. Research in these archives shows the so-called 'official' representation of Sirius, telling a story shaped by the socialist cultural policy and the current preservation practices.

The group's marginalized position in Hungary has been lack of its documentation, both musical and administrative, in terms of quality as well as quantity.

In my presentation, I will demonstrate the particular stance of cultural policies towards Sirius through various public representations of the group, shaping the possibilities to research popular music via digital databases. Sirius, as a semi-underground and barely tolerated band, has received a diverse set of interpretation, mainly via the archival practices of its active audience. Besides presenting the difference between the officially available documents and the documents being created and/or created by the fans, my goal is to highlight the importance of making the results of the do-it-yourself cultural practices digitally available, as these have a role in legitimizing the group's presence as cultural heritage.



## Eveline Wandl-Vogt

Austrian Academy of Sciences,  
Vienna, Austria

e-mail: [eveline.wandl-vogt@oeaw.ac.at](mailto:eveline.wandl-vogt@oeaw.ac.at)

**Open Innovation Research Infrastructure:  
Value driven organisational designs fostering  
innovative DH on the example of Biographical  
research**

### Keywords:

Open Innovation in Science / Digital  
Transformation / Knowledge Management /  
Knowledge Society / Network Society /  
We-Culture / Biographical Research

Biographical research aims to understand an individual's life within its social context and networks or understand cultural phenomena, whereas research infrastructures aim to meet research challenges. We understand contemporary society synonymous with networked society. It consists of public, professional, and social networks, which demonstrates a dramatic shift toward a collaborative, dynamic, pluralistic social perspective rather than one that is isolated, individualistic, and static. This idea runs counter to the long-standing tradition in Humanities research, which predominantly fosters competition, individual performance, and the maintenance of an expert culture. Research infrastructure strive to realize a collaborative turn by initiating a "sharing economy" in the academic realm. This concept of sharing and cooperation in Humanities research marks the beginning of establishing data sustainability and the unlocking of the potential of "We-culture" to extend beyond the realm of academia. In this presentation, the author introduces into the design of an Open Innovation Research Infrastructure (OI-RI) demonstrated on the example of European Scientific Academies (AGATE), submitted for the H2020 eINFra 01\_2017 Design Studies Call and an initiative and movement on discovering Open Innovation in Science. It expounds upon the implementation of Open Innovation in Science extending beyond academia into the socioeconomic and sociocultural fields. On the example of European biographical research, the author exposes how working in participatory scenarios in interaction within the proposed OI-RI could look like and how the OI-RI is supporting the need of guiding digital transformation processes, by focusing on knowledge management in value driven (organisational) partnerships and by implementing rapid innovation teams. The practical implementation of the concept takes currently place in the recently founded "exploration space" at the Austrian Centre for Digital Humanities at the Austrian Academy of Sciences, Working group "methods and innovation".

-