



EU CONEXUS

12 July 2021

10 am - 4 pm

Online



PROGRAMME

EU-CONEXUS

DAY OF

DIGITAL

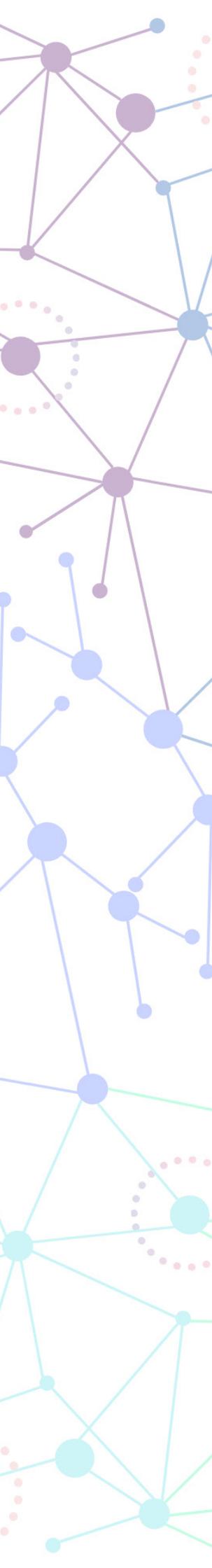
HUMANITIES



Cofinancé par le programme Erasmus+ de l'Union européenne



# PROGRAMME

- 
- 10h** Opening of the EU-CONEXUS Day of Digital Humanities
- 10h15-12h15** **Building digital tools for the Humanities (Mickaël Coustaty)**
- 10h15** *Meike Klettke (University of Rostock)* Database Backends for Digital Humanities Project
- 10h45** *Antoine Doucet (La Rochelle Université)* Robust and multilingual analysis of digitised documents - a use case with historical newspapers
- 11h15** *Holger Meyer (University of Rostock)* Graph Analysis and the Narrative Cultural Heritage
- 11h45** *Cécile Chantraine Brailon (La Rochelle Université)* L'Ecole du Spectateur : computerizing research in Performing Arts.
- 12h15-12h30** Discussion
- 14h-16h** **Building knowledge in the Humanities through digital technology (Eric Monteiro)**
- 14h** *Ludger Jansen (University of Rostock)* Ontologies: What's in it for Digital Humanities?
- 14h30** *Marijana Tomić (University of Zadar)* Glagolitic manuscript and manuscripts' fragments research in the digital environment: towards digital laboratory
- 15h** *Drahomira Cupar (University of Zadar)* The role of controlled vocabularies in the description and access to content in the digital environment
- 15h30** *Juan Gomis and Clara Bonet (Universidad Católica de Valencia)* Mapping Pliegos. A collaborative DH project on Spanish Popular Literature
- 16h-16h15** Discussion

# PARTICIPANTS & ABSTRACTS



Meike Klettke (University of Rostock)

## **Database Backends for Digital Humanities Projects**

Database backends have been developed for storing structured data, often with numerical values. The idea of a table-wise storage came up several decades ago. Storing other kind of information (more complex objects or texts) shows the limits of the relational databases.

That was the reason for the development of other data modes (XML= semistructured documents with markup). The new data models triggered the development of new database management systems for storing XML. In many application fields, workflows or graphs are a meaningful representation of the data, since some years, data models are available that can store nodes (and node labels) and edges (and edge labels) and execute standard graph algorithms. The newest format (NoSQL) offers additionally to the flexible and complex data models (JSON) also the opportunity for storing large volumes of datasets. A minor change (compared to XML) avoids their application to text formats. In the talk, a short introduction of the different database backends will be given. Two different DH sample applications: the storage and retrieval of historical (regional) land registers and the search for bias in citation graphs will be introduced.



Ludger Jansen (University of Rostock)

## **Ontologies: What's in it for Digital Humanities?**

By and by, ontologies gain ground in Digital Humanities. In this presentation, I want to explain what ontology are, what they are already successfully used for in other areas, and how Digital Humanities can

benefit from using ontologies -- namely a precise, explicit, robust and computable semantic backbone for the integration of diverse data across sources. To give some pointers for the use and development of ontologies, I present some of the results of the Good Ontology Design (GoodOD) project and discuss some best-practice rules for ontology development.



Holger Meyer (University of Rostock)

### **Graph Analysis and the Narrative Cultural Heritage**

As a computer scientist, Holger Meyer is working for twenty years in object-relational database support for digital archive and library systems.

He has a long cooperation with folklorists and ethnologists not only from the Wossidlo Archive and supervised several projects on digitizing the narrative cultural heritage of Mecklenburg. Research interests include graph management, analyzing, and mining graph data, especially on single large graphs as in social networks.

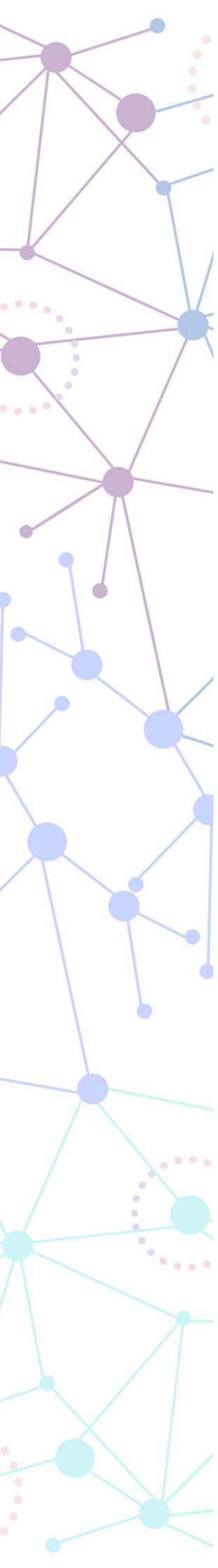


Drahomira Cupar (University of Zadar)

### **The role of controlled vocabularies in the description and access to content in the digital environment**

Subject indexing and efficient information organization are crucial in enabling access to the content. When that content is in the digital environment, usage of controlled vocabularies becomes a necessity.

Many studies showed that access enabled by usage of knowledge organization system, i. e. controlled vocabularies, such as thesaurus or subject headings system works much better than free text searching. Lately, several projects in SSH made significant effort in creation of (multilingual) thesauri. Controlled vocabularies are especially important when thinking about access to the content of the specialized (digital or digitized) collections which would be otherwise lost for the users without controlled access points. This presentation will show the example of Croatian platform Tezaurus.hr developed as a part of infrastructure for different projects under DARIAH-HR. One of the first thesaurus developed within the platform was Thesaurus for watermarks, as a part of the interdisciplinary project Digitization, bibliographic description and research of texts written on Glagolitic, Croatian Cyrillic and Latin scripts until the end of 19th century in Zadar and Šibenik area lead by Marijana Tomić, PhD from University of Zadar, Croatia. This presentation will especially focus on the methodology for development of a thesaurus using different approaches.



Marijana Tomić (University of Zadar)

### **Glagolitic manuscript and manuscripts' fragments research in the digital environment: towards digital laboratory**

This presentation will give a short overview of the interdisciplinary project Digitization, bibliographic description and research of texts written on Glagolitic, Croatian Cyrillic and Latin scripts until the end of 19th century in Zadar and Šibenik area (further Written heritage) conducted by the Centre for Research in Glagolitic of University of Zadar.

Croatian written heritage has a rich corpus of manuscripts and early prints written on three scripts – Glagolitic, Latin and Croatian Cyrillic (called bosanica) scripts, and in three languages – Church Slavonic, Croatian redaction of the Church Slavonic and Latin. The important documents from Mediaeval times onwards, namely diplomatic material, registers, fraternity documents, liturgical and literal works, etc., especially in Zadar island and hinterland, are written on Glagolitic script. Being scattered across many institutions in Croatia and abroad, either not catalogued at all, or catalogued only in the local card catalogues, those manuscripts are often hard to access, and sometimes completely hidden from researchers. The aim of the project Written heritage is to build the digital collection of high-quality digital representations of Glagolitic documents (manuscripts, watermarks, epigraphic monuments including in-situ inscriptions on stones, walls and buildings, inscriptions on clothes, objects, etc.), and to build and integrate databases with structured metadata based on controlled vocabularies and built on the international principles allowing interoperability, sharing and further research, namely visualizations of basic patterns of Glagolitics in Zadar island and hinterland. The main principle of the project is to give the platform and promote interdisciplinary research, sharing of data results and promoting research of Glagolitic within wider research community. Therefore, one of the project goals is to coordinate the interdisciplinary, cross-institutional research of Glagolitics by providing Virtual Research Environment (VRE) which is planned within GlagoLab: Portal and digital laboratory for collaborative research and promotion of the Croatian Glagolitic alphabet. Tools for cooperative transcription and annotation of manuscripts' texts, as well as crowdsourcing transcription of Glagolitic manuscripts within Civil science initiative, and tools for visualization of (meta)data will be outlined. Also, parts of the planned project concerning Glagolitic manuscripts' fragments description and reunification of original codicological units will be shown, together with IIF framework and its usage in storytelling and research of Glagolitic fragments.



Juan Gomis and Clara Bonet (Universidad Católica de Valencia)

### **Mapping Pliegos. A collaborative DH project on Spanish Popular Literature**

The web Mapping Pliegos (Mapping Pliegos (csic.es)) is the result of 5 years of collaborative work among scholars, librarians, computer scientists and several archives and libraries which hold collections of «pliegos sueltos» or «pliegos de cordel», that is, popular prints spread through Spain between the 15th and the XX centuries. Its aim is to provide the researchers and common users with a database and digital library of «literatura de cordel». In this contribution we'll first develop a general overview of this rich cultural phenomenon, which in the last years has increasingly been studied in its European dimension. Then we'll offer a description of the project: its origins and development, its current state and its future prospects and challenges.

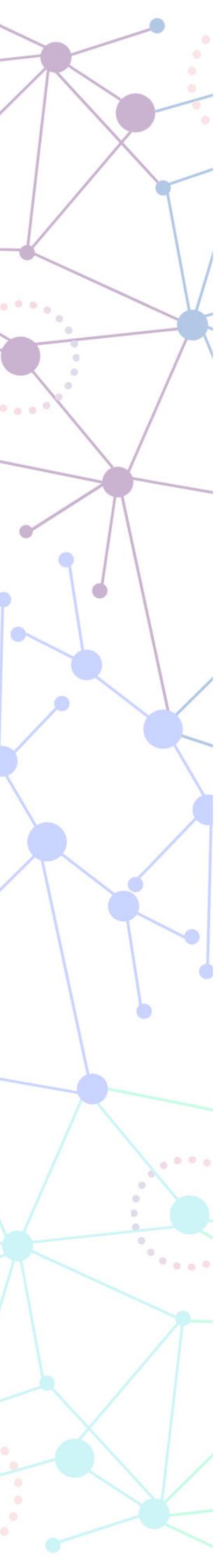


Antoine Doucet (La Rochelle Université)

### **Robust and multilingual analysis of digitised documents - a use case with historical newspapers**

Many documents can only be made accessible in the form of digitized images. This is particularly the case for historical

and handwritten documents, but for many digitally-born documents as well, turned into images for various reasons (e.g., a file conversion or a passage through an analog form in order to insert a manual signature, to send by post, etc.). Being able to analyze the textual content of such digitized documents requires a phase of conversion from the captured image to a textual representation, a key part of which is optical character recognition (OCR). The resulting text is often imperfect, to an extent which is notably correlated with the quality of the initial medium (which may be stained, folded, aged, etc.) and with the quality of the image taken from it. I will present recent advances in AI and automatic language processing enabling this type of corpus to be analyzed in a way that is robust to digitization errors. For example, I will show how we were able, in the H2020 NewsEye project to create state-of-the-art results for the cross-lingual recognition and disambiguation of named entities (names of people, places, and organizations) in large corpora of historical newspapers written in 4 languages, written between 1850 and 1950. This type of result paves the way to a large-scale analysis of digitised documents, notably able to cross linguistic borders, a key challenge in the European area.



Cécile Chantraine Brailon (La Rochelle Université)

**L'École du Spectateur : computerizing research in Performing Arts.**

L'École du Spectateur de Nouvelle Aquitaine (ESNA) is a scientific project in Digital Humanities and in Performing Arts which aims to promote research in the field using digital technology. This school will be both a school of encounters (between researchers, artists and the public) and a digital school (accompanied by the manipulation of digital tools thanks to a computer platform). Both in vivo and virtual, it can thus respond to the dual challenge of the

integral and systemic participation of digital technology in research and training in the performing arts. On the one hand, the aim is to advance the study of performative works which are by their very nature ephemeral and therefore not reproducible in an identical manner, and on the other hand, to make their production and documentation easily accessible and analysable through digital technology.