DARIAH Annual Event 2019: Humanities Data

Book of abstracts

Programme committee

Jennifer Edmond (Chair)

Femmy Admiraal (Co-Chair)

Jakub Szprot (Local organisers)

Aneta Daszuta (Local organisers)

Arianna Betti

Agiatis Benardou

René van Horik

Sarah Kenderdine

Martin Lhoták

Francesca Morselli

Tomasz Parkola

Andrea Scharnhorst

Erzsébet Tóth-Czifra

Table of contents

Workshop   27

Poster session   40

Openness and access to data   **96**

Digital collections and databases   **104**

Infrastructures and beyond   **116**

WG Meeting

Digital Practices for the Study of Urban Heritage WG Meeting

Georgios Artopoulos * [1]

[1] The cyprus institute – Cyprus

The UDigiSH Working Group focuses on the study of Digital Methods and Good Practices of heritage and urban change, the impact of urban development on cultural heritage, the identity of the city and the role of civil society. This is planned to be the first physical meeting of the participants of the WG Digital Practices for the Study of Urban Heritage. Meeting in person will accelerate interaction between the members of the group and will allow us to finalise a plan of action for the first year, the list of tasks for which will be compiled before arriving at the venue through a series of virtual meetings of the group on Zoom. The meeting will also enable us to agree on members' contribution based on availability and expertise. Coordination of activities at a local level, access to resources and sharing of tools will also be discussed during the meeting with the aim to facilitate and enable work for achieving the goals and objectives of the WG. Sub-group follow up meetings will be organised individually by various members to conduct on-site data collection/working sessions, which will be then integrated into the WG body of work. Finally, this 1st meeting of the WG will give us the opportunity to introduce our activities, motivation and objectives to other DARIAH WGs and communities, as well as to familiarise ourselves with the parallel activities of the network and explore new collaborations and opportunities that will strengthen DARIAH communities and the network.

Keywords: Digital Humanities for the Study of Historic Cities, GeoVisualisation of Urban Histories, Participatory Digital Narratives, co, creation and co, management of informal stories and memories

---

*Speaker

WG Meeting #dariahTeach

Susan Schreibman * 1

1 Maastricht University – Netherlands

This is a WG meeting for #dariahTeach. We welcome anybody who is thinking of contributing a course or workshop to the platform to attend. We will also unveil plans of transforming #dariahTeach from a project to a peer-review publicaiton and would welcome the community's feedback and ideas about this.

Keywords: teaching, education, publication, e, lerning

---

*Speaker

WG DH Course Registry

Tanja Wissik * [1], Hendrik Schmeer * [2]

[1] Austrian Academy of Sciences (ÖAW) – Austria
[2] CLARIN ERIC – Netherlands

WG DESCRIPTION

The CLARIN/DARIAH DH Course Registry offers a map-based search environment that gives access to a database that contains information on Digital Humanities courses. It currently covers only (a selection of) DH courses offered by European academic organisations, but an extension of the Registry beyond Europe may be considered later. Students, lecturers and researchers can search the database on the basis of topographical information (location), ECTS credits or the academic degrees that are awarded. In addition it is possible to search for courses based on TaDiRAH, a taxonomy of Digital Research Activities in the Humanities (including labels for tasks, techniques and objects) and sub-disciplines from the Social Sciences and the Humanities (SSH).

The goal of the DH Course Registry is to provide information to: (i) students and researchers who intend take up a study in the field of Digital Humanities, (ii) lecturers who are looking for examples of good practices in the DH field or want to promote their own DH-related teaching activities and material, and (iii) administrators who aim to attract and facilitate international student and staff mobility and exchange.

WG MEETING AGENDA

- Introduction of the working group and its goals (for new members)
- Report & updates of the DH Course Registry since last WG Meeting in May 2018
- User testing of the first version of the new interface developed within the DHCR Sustain project funded by the DARIAH Theme Call 2018
- Open discussion

Keywords: teaching, training, usertesting

---

*Speaker

Ethics and Legality in the Digital Arts and Humanities (ELDAH) WG Meeting

Walter Scholger [*] [1,2], Koraljka Kuzman Slogar [*] [3,4]

Vanessa Hannesschläger [*] [5]

[1] University of Graz (Uni Graz) – 8010, Universitätspl. 3, 8010 Graz, Austria
[2] University of Graz – Austria
[3] Institute of Ethnology and Folklore Research (IEF) – Šubićeva 42, 10000 Zagreb, Croatia
[4] Institute of Ethnology and Folklore Research – Croatia
[5] Osterreichische Akademie der Wissenschaften – Austria

DARIAH-EU promotes open access of methods, data and tools. But it also stands for a responsible conduct. As an infrastructure for Arts and Humanities researchers it will not be able to solve all issues arising in the quickly progressing domain of digital research. But with its new ethics and legal strategy it will contribute generally and especially wherever it concerns researchers from communities connected to DARIAH. This will lead to better access to tools and data for researchers as well as to guidelines and orientation for the conduct of their research. Policy makers will equally profit from the outputs of the working group as they will be facilitated with background information on issues related to the working group's topic.

The ELDAH WG has already produced material and held several workshops regarding legal issues and open licensing and intends to switch its focus towards ethical questions at this Working Group meeting.

Keywords: Ethics, Privacy, Open Science, Licensing, Copyright

---

[*]Speaker

Working Group on Visual Media and Interactivity

Wieneke Lars * [1], Adeline Joffres * [2]

[1] University of Luxembourg – Luxembourg
[2] Huma-Num : la TGIR des humanités numériques – Aix Marseille Université : UMS3598, Centre National de la Recherche Scientifique : UMS3598 – France

Visual media and interactivity gain more and more traction in the domain of digital humanities. Therefore, the DARIAH WG Visual Media and Interactivity in its second revision aims at becoming a hub for all activities related to this topic with the idea to connect people, projects and ideas from within DARIAH and beyond.

We are primarily interested in creating synergies between ongoing and future research projects and to guide interested researchers in the application of tools, services and methods from the visual media and interactivity domain as well as to discover what DARIAH has to offer. We also understand ourselves as bridge builders between digital humanities and the wider field of cultural heritage computing.

We want to use this meeting to

a. take stock in the interest of the community in visual media and interactivity
b. to recruit new members that will help us to fill the initiative with life as well as
c. find a new co-chair for our WG.

Keywords: visual media, interactivity, applications and tools, research, cultural heritage, cultural heritage computing

---

*Speaker

WG meeting Sustainable publishing of metadata

Dorien Styven [*] [1], Johan Van Der Eycken [*] [2]

Tom Gheldof [*] [3]

[1] Kazerne Dossin (KD) – Goswin de Stassartstraat 153, 2800 Mechelen, Belgium
[2] State Archives – Belgium
[3] KULeuven – Belgium

The annual WG meeting will this year focus on the publication of the papers presented at the workshop in Brussels in May 2018.

_____
[*]Speaker

Digital Methods and Practices Observatory (DiMPO) WG meeting I: scholarly practices and DH meta-research showcase

Costis Dallas * [1,2], Maciej Maryl * [3]

[1] Faculty of Information, University of Toronto (iSchool, UofT) – 140, St George street Toronto ON M5S 3G6, Canada
[2] Digital Curation Unit, IMIS/Athena Research Centre (DCU-ARC) – Artemidos 6 Epidavrou GR-151 25 Maroussi, Greece
[3] Institute of Literary Research, Polish Academy of Sciences – Poland

The purpose of this meeting is to showcase and discuss recent work by DiMPO WG and related initiatives to study digitally-enabled scholarly practices through the application of qualitative research methods, including the DH meta-research methodological approach developed by DiMPO members in the context of the Digital Humanities Work in Focus (DH-FOCUS) DARIAH CIO grant last year. Additional discipline-specific initiatives conducted either within trans-European COST networks or initiated by researchers in countries such as Finland and Switzerland will also be presented.

The meeting is open to everyone interested to learn more on scholarly practices research for the DH community, and to find how they can be part of a coordinated multicase studies project planned for the forthcoming period by DiMPO WG. Colleagues who wish to present their work on qualitative research on scholarly practices in the arts and humanities, including but not limited to interviewing, case studies and ethnographic inquiry, are warmly invited to contact the session organizers, Costis Dallas (c.dallas@dcu.gr) and/or Maciej Maryl (maciej.maryl@ibl.waw.pl) with a short message introducing their proposed topic.

The meeting will be based on a sequence of short (10 minute) presentations, followed by discussion. A preliminary programme (as of 6 March 2019) follows:

- Case study protocol for meta-research in DH prepared under DH in Focus DARIAH Grant (Costis Dallas)

- Research community profiles and qualitative content analysis under COST Action NEP4DISSENT (Jennifer Edmond, Jessie Labov, Costis Dallas, Ingrida Kelpšienė, Maciej Maryl)

- Episodic interviewing of Polish scholars in the humanities (Marta Kolodziejska, Klaudia Grabowska)

---

*Speaker

- Qualitative content analysis of focus groups of archaeology-related communities under COST Action ARKWORK (Ingrida Kelpšienė, Costis Dallas)

- A panorama of situated research in the digital humanities: the 'Lab & Slack' DHQ special issue (Mila Oiva)

- A pilot mixed methods study of theology and religious studies scholarly practice in Switzerland (Sara Schulthess, Claire Clivaz)

Further presentations: please contact session organizers to be considered.

Keywords: research methods, multicase studies, survey, DiMPO, DH meta, research, qualitative research, scholarly practices

Women Writers in History : WG meeting

Amelia Sanz * [1], Suzan Van Dijk * [2]

Alicia Montoya * [3]

Beatrjis Vanacker * [4]

Viola Parente-Čapková * [5]

[1] Complutense University of Madrid – Spain
[2] Huygens Institut – Netherlands
[3] Radboud university [Nijmegen] – Netherlands
[4] KULeuven – Belgium
[5] University of Turku – Finland

In this internal WWIH-WG meeting, we will monitor activities developed all over the year and we will plan our agenda for the second half of 2019 and 2020. Some other WG members will participate by videoconference

Keywords: women writers, database, teaching

*Speaker

A web of thesauri: Aligning specialist thesauri to the Backbone Thesaurus.Thesaurus Maintenance WG meeting

Helen Goulis * [1], Eleni Tsoulouha * [2]

[1] Academy of Athens – Greece
[2] FORTH-ICS – Greece

Main organisers: Helen Goulis (Academy of Athens) co-chair, Eleni Tsouloucha (FORTH-ICS) co-chair

Established in the context of DARIAH-EU, the Thesaurus Maintenance WG (TMWG) aims at developing a model to render existing terminologies used across the humanities, interoperable and effectively managed. The compilation of a common and consistent overarching thesaurus for the humanities (Backbone Thesaurus-BBT), under which all the vocabularies used in the different disciplines can be mapped to one set of top level terms, lies in the core of this effort. To this end, we focus on uncovering the semantic principles and the methodology guiding the selection of the concepts that serve as a common ground for thesaurus building –and, in that sense, enable cross-disciplinary resource discovery.

BBT runs on BBTalk, a platform facilitating the dynamic alignment of specialist thesauri to the BBT –in the sense that the users of BBTalk can make requests for changes on the structure of the BBT, to the end that it accommodates thesauri covering highly specific and developing areas of knowledge. BBTalk is currently used by the BBT managing team; a group of scholars in charge of discussing and implementing the proposed changes on BBT, whose members are affiliated with DARIAH and the TMWG. In this session we will be exploring topics regarding the methods to building and managing thesauri, exemplified by use cases from the BBT, and the specialist thesauri of the Academy of Athens (AA), the Deutsches Archäologisches Institut (DAI), and the Fédération et ressources sur l'Antiquité (FRANTIQ).

TENTATIVE AGENDA:

- Report on the activities of the WG over the last year, building up to the presentation of the BackBone Thesaurus (BBT) –scope of the endeavor, structure of the BBT and methods followed in its making–, the tools developed to enable managing the BBT (BTalk).

---

*Speaker

- Use cases of alignment of local thesauri to the BBT.

- Exchange on how to structure specialist thesauri. Educational/training material on thesaurus-building.

- How to increase the impact of our work/joint projects, leading to greater dissemination and new collaborations (open invitation to thesauri holders).

- Side-projects that might be of interest to the TMWG. Discussion of the terms and hierarchies of the BBT, questions regarding the alignment of specialist thesauri to the BBT – depending on participation.

We wish to bring a poster for the Marketplace.

List of envisioned participants

Gerasimos Chrysovitsanos (Academy of Athens)

Patricia Kalafata (Academy of Athens)

Blandine Nouvel (Centre national de la recherche scientifique)

Evelyne Sinigaglia (Centre national de la recherche scientifique)

Matej Durco (Österreichische Akademie der Wissenschaften)

David Wigg-Wolf (German Archaeological Institute)

Others to be confirmed

Keywords: Thesauri, Ontologies, Interoperability, Terminology systems, Controlled vocabularies, Interdisciplinarity, Semantic structuring, Semantic linking

Working Group Meeting GiST - What does the Standardization Survival Kit (SSK) tells us about Humanities Data?

Klaus Illmayer * [1], Marie Puren * [2]

Charles Riondet * [3]

[1] Austrian Academy of Sciences (OeAW) – Doktor-Ignaz-Seipel-Platz 2, 1010 Wien, Austria
[2] CNRS – UMR LARHRA – France
[3] Inria-EPFL – Institut National de Recherche en Informatique et en Automatique – France

First of all we like to give an introduction to the Standardization Survival Kit (SSK)[1]. This will be an open session, where everyone can join. We will explain the mission of the SSK with a focus on research scenarios that deal with humanities data (textual resources, archaeological data, digitization processes, etc.). The SSK is a way to promote the use of standards and supports researchers to find documentation, guidelines, and examples for digital research use cases. The SSK can be found online at http://ssk.huma-num.fr, its documentation is available at https://ssk.readthedocs.io.
We also like to give some insights into the already existing scenarios of the SSK and how humanities data is represented there, by presenting the current status of the SSK scenarios landscape and focusing on examples relevant for the DARIAH communities. We expect some discussion on this, so that the WG GiST can integrate feedback in the working plan.

The second part of the meeting is about shaping the working plan of the WG GiST and discussing the integration of the SSK in the activities of the WG GiST. This will also cover sustainability, curation, and expanding of the SSK, in particular by the creation of a dynamic users' community. We also like to discuss the integration of the SSK in the SSHOC marketplace. People who like to join the WG are very welcome.

Charles Riondet also proposed a poster on the SSK. It would be great If our WG meeting is after the poster presentation, so that we can directly invite interested people to join the WG meeting at least for the first part.

Agenda

(for a 2 hours meeting)

---

*Speaker

0:00 - 1:00 Introduction of the SSK and discussion on the representation of humanities data in the SSK

1:00 - 1:30 Integration of the SSK in the WG GiST

1:30 - 2:00 Discussing and finalizing working plan of the WG GiST for 2019/2020

List of participants

WG chairs: Klaus Illmayer, Marie Puren, Charles Riondet

VCC3 head: Tomasz Parkola

WG members: we expect that at least two other members of the WG are present

We also expect that some interested people join at least the introduction of the SSK. There will be also a poster on the SSK, so maybe it makes sense that the WG meeting is after the poster presentation, so that we can directly invite interested people to join the WG meeting. Everyone is also invited to join the second part of the WG meeting with more internal topics to discuss. It would be great if the meeting helps us in engaging new members for the WG.

—

The Standardization Survival Kit (SSK) is an open tool, developed by the PARTHENOS project and supported by the WG GiST, that helps researchers choose standards and best practices for the application of digital methods. Build up on the idea of providing research scenarios, it establishes a low-barrier entry point to get an overview on used standards in different research fields.

Keywords: standards guidelines

1st DARIAH-AIM Meeting: The Artificial Intelligence and Music WG at the DARIAH 2019 Annual Event

Albert Meroño Peñuela * [1], Enrico Daga * [2]

[1] Vrije Universiteit Amsterdam [Amsterdam] – Netherlands
[2] The Open University [Milton Keynes] – United Kingdom

Organizers/WG chairs. Albert Meroño-Peñuela and Enrico Daga

Meeting description. The WG on AI and Music (AIM) has spent its first months in starting its advocacy efforts, gathering a community around AI and Music, and mapping the various fields working on the subject. This is going to be the first face to face meeting of the core WG members and its most immediate surrounding community. The objective of AIM is to strengthen all fronts of advocacy at the intersection of Music and Artificial Intelligence. In order to do this, AIM addresses the outreaching needs of universities, industry and institutions by creating a map of communities around AI and Music and providing them with various communication and exchange means. Therefore, the goal of the meeting is to summarize the state of affairs regarding this map and communication means, and establish an agenda for the coming year.

Agenda. The provisional agenda for the meeting is as follows:

State of the WG (presentation by chairs) [20 minutes]

Community map document [40 minutes]

Coverage of communities in the mailing list

Communication and exchange [40 minutes]

Current facilities

Webinar schedule

Ties with other DARIAH WGs

Suggestions, comments, remarks [20 minutes]

---

*Speaker

Envisioned participants. The list of expected participants is as follows:

Albert Meroño-Peñuela, Vrije Universiteit Amsterdam

Enrico Daga, The Open University

Aldo Gangemi, University of Bologna

Mathieu d'Aquin, Insight Centre for Data Analytics

Robin Laney, The Open University

Simon Holland, The Open University

Johan Oomen, Netherlands Institute for Sound and Vision

Gregory Markus, Netherlands Institute for Sound and Vision

Martin Clancy, Trinity College Dublin

Poland AI and Music group

David de Roure, University of Oxford

Kevin Page, University of Oxford

Pasquale Lisena, EURECOM

Daniel Bangert, Göttingen State and University Library

"Bibliographical Data" Working Group - Current State of the Initiative

Vojtěch Malínek * [1], Tomasz Umerle * [2]

[1] Institute of Czech Literature, Czech Academy of Sciences – Czech Republic
[2] The Institute of Literary Research of The Polish Academy of Sciences – Poland

Comment: Although the group has not been formally established yet, DARIAH-EU management was contacted (Francesca Morselli) - through Maciej Maryl, one of the future members -, and gave the green light to the initiative, and included a mission statement of the proposed Group in the official newsletter.

We would love to take advantage of the DARIAH event to meet with potential members, and VCC3 co-chair, and discuss the current state of affairs. The fact that the meeting is in Warsaw is also an important factor, as current coordinators are from Poland and Czech Republic, and some other future members will attend the event.

In recent weeks we have been working on soliciting interest in the WG, and are nearing the end of a series of talks with interested members (list below). In April we will focus on working on the documentation, and after that send it to the DARIAH-EU management.

Note: also an informal meeting of the members can be considered during these days, if this application is rejected, but we decided to apply for the inclusion of the meeting in the programme, as it might serve as an additional opportunity to garner further interest in the WG.

Main Organizers:

Vojtěch Malínek, Tomasz Umerle

Main topic of the meeting:

Current state of the works toward establishing a WG: final version of mission statement, confirmed members, and main goals for the first year of WG activity, especially in the context of VCC3.

Preliminary agenda:

presentation of the final version of the mission statement consulted with all the members;

---

*Speaker

short presentations of the members and their research interests/projects;

meet-up with Tomasz Parkola, co-chair of VCC3;

main goals for the first year of WG's activity.

Envisioned Participants (some will be present online, for example via Skype, but at least 8 declared to be present in person):

Tomasz Parkola, Poznań Supercomputing and Networking Centre, Co-Chair of DARIAH's VCC3, Poland

Vojtěch Malínek, Czech Literary Bibliography, Institute of Czech Literature, Czech Academy of Sciences, Czech Republic

Tomasz Umerle, Polish Literary Bibliography, Institute for Literary Research of the Polish Academy of Sciences, Poland

Alexandre Guilarte, Bibliography of Irish Linguistics and Literature, School of Celtic Studies, Dublin Institute for Advanced Studies, Ireland

Václava Horčáková, Bibliography of the History of the Czech Lands, Institute of History, Czech Academy of Sciences, Czech Republic

Peter Kiraly, Gesellschaft für wissenschaftliche Datenverarbeitung, University of Göttingen and Max Planck Society, Germany

Leo Lahti, Helsinki Centre for Digital Humanities, University of Turku, Finland

Maciej Maryl, Digital Humanities Centre, Institute for Literary Research of the Polish Academy of Sciences, Poland

Volker Michel, Bibliography of German Language and Literature, Goethe University Frankfurt am Main, Germany

Alicia Montoya, Radboud University, Netherlands

Róbert Péter, Institute of English and American Studies, University of Szeged, Hungary

Allejandra Ulla Lorenzo, Universidad Internacional de La Rioja, Spain

Adam Pawlowski, University of Wroclaw, Poland

Mikko Tolonen, Helsinki Centre for Digital Humanities, University of Helsinki, Finland

Piotr Wciślik, Digital Humanities Centre, Institute for Literary Research of the Polish Academy of Sciences, Poland

Poster for the Marketplace:

Yes. Poster will briefly describe main goals and expected activities of the proposed working group. It shall mainly introduce WG and its plans for the future to the interested audience.

Digital Methods and Practices Observatory (DiMPO) WG meeting II: reports on past work and workplan (2019-20)

Costis Dallas * [1,2], Maciej Maryl * [3]

[1] Faculty of Information, University of Toronto (iSchool, UofT) – 140, St George street Toronto ON M5S 3G6, Canada
[2] Digital Curation Unit, IMIS/Athena Research Centre (DCU-ARC) – Artemidos 6 Epidavrou GR-151 25 Maroussi, Greece
[3] Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences – Poland

This is the business meeting of the DiMPO Working Group. Its objectives are review past WG activities, examine the current situation and needs on the scope of the WG, and on this basis consider and decide on the WG workplan for the forthcoming year.
Agenda:

1. Introductions and welcoming new WG members

2. European survey on scholarly practices and digital needs - official publication of full report (Costis Dallas)

3. Revisions to survey questionnaire and planning for next iteration of survey (Costis Dallas & Maciej Maryl)

4. Protocol and planning for multiple case studies research (Maciej Maryl and Costis Dallas)

5. Zotero bibliography on DH practices and methods (Orla Murphy)

6. Proposals for new WG projects/activities

6. Reports on member activities - any other business
DiMPO is open to discuss proposals for new projects within the scope of the WG, beyond currently active projects (survey, multiple case studies, Zotero bibliography). To ensure good planning, please share ideas for new projects with c.dallas@dcu.gr and maciej.maryl@ibl.waw.pl before the meeting.

---

*Speaker

Keywords: survey, multicase studies, Zotero bibliography, business meeting

Workshop

Swimming in (Linked) Data: Using the Semantic Web to make sense of Spatial Humanities

Rebecca Kahn * [1], Piraye Hacigüzeller * [2,3]

Patricia Murrieta-Flores [4]

[1] Alexander von Humboldt Institute for Internet and Society (HIIG) – Französische Straße 9 10117 Berlin, Germany
[2] University of Oxford – United Kingdom
[3] Univeristy of Leuven – Belgium
[4] Lancaster University – Bailrigg, Lancaster. UK LA1 4YW, United Kingdom

The availability of born digital data as well as digitised collections, is changing the way we study and understand the humanities. This amount of information has even greater potential for research when semantic links can be established, and relationships between entities highlighted. The work of several projects presented in the workshop have shown that connecting historical data according to their common reference to places (expressed via URIs stored in gazetteers) is a particularly powerful approach to investigating digital textual and image-based sources. The power of linked data is that it enables search and query beyond one's own projects, and allows leverage of the quantity of relationships between semantic connections that can be made and discovered.

This workshop is designed to show that working with semantic data searches is not meant to render specific results, but rather to highlight how, just as humanities scholarship is about interpretation, association and occasional serendipity, digital humanities scholarship can be used to discover distributions, anomalies, or patterns, rather than specific answers.

This workshop will introduce a range of tools which, when used as part of a pipeline, can enable the extraction, linking and visualisation of semantic data found in historical sources. These will include tools for creating annotations (Recogito) conceptualising linked open data (a TripleStore) and sharing linked data (Wikidata and Peripleo).
Recogito an award winning tool developed by Pelagios that enables annotation of place references in text, images and data through a user-friendly online platform. The principal function of Recogito is the ability to allow non-experts to produce semantic annotations, while at the same time allowing the user to export the data produced as valid RDF, XML, GeoJSON and TEI formats. Participants will be able to explore Recogito's newer features, such as collaborative annotation and controlled vocabularies, as they upload and annotate text and images and then download their annotations in the available data formats. By exploring sources and discussing the related challenges of using a wide range of sources, participants will be able to discuss how

---

*Speaker

Recogito might be used to support their own research.

Attendees will also be invited to explore Peripleo, a map-based visualisation for exploring Linked Open Data relationships. Working with over 8 million objects, across almost 2 million years, users will be able to search across a linked data ecosystem for places, documents, objects and keywords.

Wikidata is increasingly being used by humanities and GLAM researchers as both a source of linked open data, and as a mechanism for connecting other collections of data (in particular GLAM sources) to each other. This workshop will give a brief introduction to how researchers can contribute their data to Wikidata, thus adding to the web of Linked Open Data, and also use it as an interconnection format, to create hybrid data sources.

We would recommend running these explorations across 2 workshop sessions, in order to give attendees enough time to get to grips with the tools. While there will be short introductions and time for discussion, most of the time will be used for hands-on, practical experimentation.

This workshop would also serve to form part of the stated activities of Pelagios' 2018 DARIAH Theme grants agreement, which aims to establish the use of Linked Data within scholarly practice by establishing its use within the DARIAH ecosystem of tools and services, and to bring DARIAH community members into the broader LOD community by formalising existing, informal collaborations with DARIAH Working Groups.

Keywords: Geo, humanities, spatial humanities, linked data, annotations, Wikidata

Ensuring the sustainability of the Open Marketplace: a community-driven approach

Frank Fischer * [2,1], Laure Barbot * [2]

[2] Digital Research Infrastructure for the Arts and Humanities – Germany
[1] Higher School of Economics – Russia

Building and sustaining the Marketplace to facilitate fluid exchange of tools, services, data and knowledge is one of the four pillars of the DARIAH strategy for the upcoming years. Since last year, we know that this objective will be undertaken as Work Package 7 (WP7) in the framework of the Social Sciences and Humanities Open Cloud (SSHOC) project, a cluster project run by several SSH infrastructures and partners seeking to implement services into the European Open Science Cloud (EOSC) for their respective communities.

As part of a long-term process (and not just an output of a work package in a three-year project) and to ensure its own sustainability, the Marketplace can not start from the scratch. The project is already taking into account results of work previously done in a DARIAH context (WGs surveys, CENDARI, Humanities at Scale, PARTHENOS, DESIR...) and beyond (TAPoR, Bamboo project and DIRT Directory), and one of the objective of the Marketplace is to achieve a convergence with existing solutions and services, and that's why it is inherently a community-driven project.

The objective of this 2-hour workshop during the DARIAH annual event is to involve DARIAH's community in the first steps of the definition/specification of the Marketplace, especially to address the integration of Humanities data and methods. In that sense, this workshop also answers to action 1 of the Strategic Action Plan II ("DARIAH must develop and deploy mechanisms to ensure that the Marketplace vision and value is embedded in and driven by our community, and has a larger place in organisational thinking and planning than the SSHOC project alone.").

After a short presentation of the Marketplace in the SSHOC project context and the envisioned steps to involve the different communities, we would like to open a discussion on the two following topics:

What's in the Marketplace for you? As DARIAH's members and stakeholders of the different projects that have been involved in the design of the Marketplace idea, what do you envision for the Marketplace?

Beyond the general expectations regarding the Marketplace, this question could be an opportu-

---

*Speaker

nity to work on user stories based on real researchers' needs. These user stories could be added to the Marketplace user requirements that DARIAH will have to deliver for the SSHOC project.

How can we ensure that the Marketplace will be inclusive and community-driven?

Several corresponding tasks are already planned within the SSHOC project, but we could/should add a DARIAH component:

create a set of materials that national partners could use during their events

involve DARIAH Working Groups in the test of the different versions of the Marketplace

organise a survey within DARIAH community (when? who? on which aspects of the MP?)

organise a Guerilla testing (when? who?)

Expected results

The outputs of this workshop should be reused as part of the community consultation process that will be described in the first WP7 deliverables of the SSHOC project but not only. As this workshop aims also to obtain input and feedback from DARIAH stakeholders regarding the participative method itself, the goal is also to precise this methodology and to integrate new ideas of the participants.

Structure

1. Introduction (DARIAH/SSHOC/Marketplace) - 5 min

2. General presentation of the SSHOC project - 10 min

2.a. Implementation timeline for the Marketplace with a focus on participatory phases

2.b. Recap of work already done and links with previous projects

3. Presentations about the involvement of the communities on the Marketplace:

3.a Relation between WP6 and WP7: interlinking and discovery of tools, solutions and training material - 5 min

3.b Curation and governance of the Marketplace - 5 min

3.c Participatory approaches (user experience and participatory design, user stories mapping) - 5 min

4. Discussion based on two main questions: consultation on the Marketplace contents and consultation on the participative methodology - 1h15

5. Conclusion and next steps - 15 min

List of speakers

Matej Durco (OEAW). 3.a: Relation between WP6 and WP7: interlinking and discovery of tools, solutions and training material.

Suzanne Dumouchel & Clara Petitfils (CNRS Huma-Num). 3.b: links with the task 7.4 of the SSHOC project "Governance: Population, Curation & Sustainability of the Marketplace".

Frank Fischer & Laure Barbot (DARIAH-EU). General Presentation

Dorian Sellier *TBC* (INRIA) & Stefan Buddenbohm (UGOE - SUB). 3.c: how UX and participatory approaches can be an answer to the community involvement?

A CLARIN representative

Keywords: Open Marketplace, community, sustainability, SSHOC, EOSC

Ethical and legal considerations in obtaining, sharing and reusing data from human participants - workshop proposal

Erzsébet Tóth-Czifra * [1], Koraljka Kuzman Slogar * [2,3]

Walter Scholger * [4]

[1] Digital Research Infrastructure for the Arts and Humanities – Germany
[2] Institute of Ethnology and Folklore Research (IEF) – Šubićeva 42, 10000 Zagreb, Croatia
[3] Institute of Ethnology and Folklore Research – Croatia
[4] University of Graz (Uni Graz) – 8010, Universitätspl. 3, 8010 Graz, Austria

It is now beyond question that opening up access to scholarly knowledge is a key value of the academy in the 21st century. The Open Science paradigm however is not only restricted to facilitating democratic access to knowledge but also accommodates a set of key values of present-day knowledge creation such as fairness, transparency, equality and increased rigour and accountability in scholarly activities. Open Science therefore involves a cultural shift in the science ecosystem where research workflows are made transparent and subject to greater scrutiny from the early stages of protocols and planning through publication and data sharing. This is especially the case in research projects involving humans participants as subjects. Privacy and confidentiality become key ethical questions when research data is based on humans. For that reason, modern researchers are facing a complex set of legal issues and ethical dilemmas whenever they want to store, use, publish and share data collected from human participants. On the top of such dilemmas, researchers are also challenged by the perceived confrontations between the open research culture and the proliferation of ethical review procedures and legal requirements for data protection.

In our workshop, we are aiming to address such challenges and deliver targeted advocacy for arts and humanities scholars working with data obtained from human participants. Making good use of the knowledge base and previous work that have been carried out in DARIAH's Ethics and Legality in Digital Arts and Humanities (ELDAH) Working group, we will recommend concrete steps that should be taken in order to secure ethically responsible research and present several guidelines and regulations that exist on issues related to informed consent and confidentiality.

Outline

The workshop will be organized along the following proposed outline:

---

*Speaker

Introduction into ethical and legal issues in conducting research with human participants. How to align openness with proper data protection? (15 minutes)

What data protection is all about and what has changed with GDPR? Rules, principles, and good enough practices from case studies (30 minutes)

Presentation and analysis of GDPR-compliant consent form samples (15 mins)

Breakout session: create your own consent form. During this session workshop attendees will have the opportunity to work together on consent forms designed to different research scenarios such as organizing a conference, conducting a research project based on interview data, or designing a citizen science project and discuss the methodological pitfalls, unexpected challenges and solutions that might occur in engaging such activities. (45 minutes)

Wrap-up, takeaways, questions (15 minutes)

Planned outcomes

Our previous experience gained from workshops and collaborations of different kinds with arts and humanities research communities clearly shows that there is an increasingly strong need for advocacy in legal and ethical issues. As the main outcome of the workshop, we aim to equip arts and humanities scholars with tools, practical advice but also with twists in mindsets that will help them in establishing ethically committed, optimal research workflows. We wish to showcase and co-create research scenarios in which the open research culture is well aligned with responsible conduct and where the reuse potentials of hard-obtained data are maximized. On the other hand, the conversation with the workshop attendees and their input from their own research settings will expand the knowledge base of the Working Group as it provides us with valuable insights fuelling better-informed and better-targeted advocacy.

Speakers

Koraljka Kuzman Šlogar

Walter Scholger
Erzsébet Tóth-Czifra

Keywords: Open licensing, Research ethics, GDPR, Legal issues, Data from human participants, Humanities data, Consent forms, ELDAH

Cultural Heritage Reuse Cases: Facilitating the Exchange of Perspectives between Heritage Professionals and Researchers

Alba Irollo * [1], Erzsébet Tóth-Czifra * [2]

[1] Europeana Foundation – Netherlands
[2] DARIAH – Germany

Identifying the shared needs of the hugely diverse communities falling under the umbrella of humanities is not an easy task. Improving access to cultural heritage data can be recognized as one such widely shared needs. Since there are hardly any disciplines where source materials curated in and produced by museums, galleries, libraries and archives are not widely, or not used at all, improving access to and the reusability of these sources is one of the major data management challenges in the field of arts and humanities. The lack of a clear and comprehensive framework that could serve as a general baseline for interactions between scholars and cultural heritage institutions can be recognized as one of the greatest barriers in optimal reuse of Cultural Heritage data.

To find solutions to this problem, several European organizations such as APEF, CLARIN, DARIAH, Europeana, E-RIHS and European projects like Iperion-CH, PARTHENOS came together under the governance of DARIAH to work together and set up principles and mechanisms for improving the use and re-use of cultural heritage data issued by cultural heritage institutions and studied and enriched by researchers.

The first step of this activity concerned the definition and validation of generic principles that Cultural Heritage Institutions (CHIs) and researchers could both adhere to around the themes: Reciprocity, Interoperability, Citability, Openness, Stewardship and Trustworthiness.

To further align this vision with communities of practice, and to outline the next steps towards the implementation of the Charter, this workshop will focus on the identification of the major gaps that typically occur in CHI professional-researcher interactions under different settings.

The two hours workshop will allow us to bring together (re)use case studies from the cultural heritage sector presented by Europeana Research and from the research world presented by humanities scholars from the DARIAH network. These inputs will be co-analyzed and discussed together to uncover those critical aspects of the respective workflows where the two perspectives find less resonance with each other.

---

*Speaker

Outline

- Introduction and summary of the progress made on the Heritage Data Reuse Charter (Erzsébet Tóth-Czifra, DARIAH-EU; 10 minutes)

- Introduction about Europeana Research and the Europeana Research Community (Alba Irollo, Europeana Foundation; 10 minutes)

- Reuse cases presented by Cultural Heritage professionals (Steven Claeyssens, Curator of Digital Collections at the National Library of Netherlands, The Hague / NL); Erik Buelinckx, Head of Photo Collection, Library and Digitisation at the Royal Institute for Cultural Heritage, Brussels / BE); 2x15 minutes + 15 minutes debate

- Reuse cases presented by arts and humanities scholars (Maire Puren, the SILKNOW project; Sally Chambers The Ghent Centre for Digital Humanities); 2x15 minutes + 15 minutes debate

- Wrap-up (10 minutes)

———-

Europeana Research intends to bring a poster for the DARIAH Marketplace Poster session as well.

Keywords: Cultural Heritage, Reuse, GLAM sector, Research

Let's Be FAIR: Forging Organisational Recommendations on Research Data in the Humanities

Natalie Harrower * [1], Maciej Maryl * [2]

Eveline Wandl-Vogt * [3]

[1] Digital Repository of Ireland, Royal Irish Academy – Ireland
[2] Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences – Poland
[3] Austrian Centre for Digital Humanities, Austrian Academy of Sciences – Austria

1. Rationale

The principles of Open Science are currently debated in all areas of research. In particular, the FAIR principles are gaining ground: data must be constructed, stored, presented, and published in such a way that they can be retrieved, accessed to, reused, and made interoperable. While the implementation of these principles was first proposed in the life, natural and technological sciences, it is now clear that all that is relevant across all disciplines. The opening, communication, appropriation and reuse of research data, whatever it may be, is a common goal across research domains, but it requires a change in the understanding of data in the research lifecycle, as well as changes to research practice, and institutional research culture.

Humanities data can be rich and complex, non-standardised in format, without common or consistent metadata and ontologies, and can be subject to complex rights issues. Consensus and best practice regarding digitisation and metadata standards for common usage, that still retain the richness of different disciplines and data types, could enable open access to Humanities data, and facilitate data exchange and sharing between the wealth of archives, repositories and libraries across Europe.

The ALLEA e-humanities working group is conducting an open consultation to determine what tooks, workflows, training and resources are required by Humanities researchers. This workshop, organised in coordination with the DiMPO WG, will be focused on determining the role that research institutions need to play in the data sharing ecosystem. Participants are asked to contribute to the following questions: what do you require from your institution in order to create FAIR and open data? What resources are in place, and what are urgently required? Where

---

*Speaker

can you draw on support (policy, workflows, tools, training, etc.) in order to move institutional practices forward? etc.

The outcomes of the workshop will be incorporated into the forthcoming ALLEA e-humanities WG guidelines for European Academies on institutional practices concerning FAIR open data.

2. Workshop format

Target audience: scholars, archivists and cultural heritage professionals in various humanities disciplines from the DARIAH-community who work with data or have an interest in data.

Technical support needed: projector, post-its, markers, whiteboard/flipchart, wireless internet

What participants can gain: shared knowledge of workflows and resources for better data sharing within their institution

Contribution to the DARIAH community: recommendations will be incorporated into both DARIAH DiMPO outputs, as well as the forthcoming recommendations to Humanities researchers being prepared by the ALLEA e-humanities working group

Participation requirements: We accept applicants on a first-come, first-served basis, although in case of a waiting list each project team will have the opportunity to be represented by only one team member.

Method: World café with questions addressing various institutional aspects of research data in the humanities. The workshop will build upon the versatile experience of Austrian Centre for Digital Humanities, covering such techniques as user stories, process collage, thermometers of indicators, dotmocracy or planning poker as well as canvases, adapted for this workshop by ACDH exploration space.

Tentative outline:

a. Brief introduction to the core concepts: FAIR principles; participatory knowledge creation, design thinking and problem crowdsourcing on the given case study; ALLEA eHumanities group related output

b. Core questions, relevant to the workshop topic, are identified through the participatory process. In a world café setting, participants divided into tables are discussing and presenting their outcomes

c. In the summary phase, the core challenges are identified and addressed in more detail if the timing allows. The challenges are discussed in a guided process in breakout groups.

3. Organisers

This workshop is jointly proposed by ALLEA e-humanities working group and DARIAH DiMPO WG.

About the ALLEA e-humanities WG

The European Academies constitute a unique pan-European knowledge base that is trusted,

non-partisan and long-term. The Academies therefore have an important contribution to make to debates regarding sustained digital infrastructures and project-funded artefacts, the achievement of long-term durable digital preservation, and the societal responsibility for preservation of our digital cultural heritage. The E-Humanities Working Group is charged with identifying and raising awareness for priorities and concerns of the Digital Humanities, and contributing to the Open Access agenda from a Humanities and Social Sciences perspective, and building consensus for common standards and best practices in E-Humanities scholarship and digitisation. The Group's first publication, Going Digital: Creating Change in the Humanities, made recommendations around archival sustainability and data training required for achieving Open Access and Open Data goals across the Humanities. Going forward, the E-Humanities Working Group is focusing on the European Open Science and Open Research agendas, identifying growth opportunities for Humanities scholarship, as well as the contributions Humanities methodologies can make to truly opening research.

About the DARIAH DiMPO WG

The DARIAH Digital Methods and Practices Observatory WG (DiMPO) aims to develop and provide an evidence-based, up-to-date, and pragmatically useful account of the emerging information practices, needs and attitudes of arts and humanities researchers in the evolving European digital scholarly environment, for the benefit of the digital humanities research community. It seeks to achieve this objective through the inception of a longitudinal mixed methods research and monitoring programme on the information practices and scholarly methods employed in digitally-enabled arts and humanities work across Europe, and through the digital dissemination, validation and enrichment of research outcomes by the scholarly community. Currently, the WG is conducting a series of qualitative case-studies, preceding the next iteration of its survey. Insights into scholarly practices and institutional obstacles regarding research data in the humanities will contribute to the group's mapping of the contemporary digital scholarship.

Poster session

Ethics and Legality in the Digital Arts and Humanities (ELDAH)

Walter Scholger * [1], Koraljka Kuzman Šlogar * [2]

Vanessa Hannesschläger * [3]

[1] University of Graz (Uni Graz) – 8010, Universitätspl. 3, 8010 Graz, Austria
[2] Institute of Ethnology and Folklore Research (IEF) – Šubićeva 42, 10000 Zagreb, Croatia
[3] Osterreichische Akademie der Wissenschaften – Austria

The European Research Infrastructure Consortium "Digital Research Infrastructure for the Arts and Humanities" (DARIAH-EU) promotes open access to methods, data, and tools, and stands for a responsible conduct. As an infrastructure for Arts and Humanities researchers, it strives to deal with all issues arising in the quickly progressing domain of digital research. DARIAH's new ethics and legal strategy - e.g. the establishment of a dedicated Open Science Officer and the Working Group on Ethics and Legality in Digital Arts and Humanities (ELDAH) - is an effort to establish openness as a de-facto standard in digital research and a contribution to ensure a high standard of legal and ethical awareness wherever researchers from communities connected to DARIAH are concerned. This should facilitate better access to tools and data for researchers, as well as to recommendations and best practices for the conduct of their research. The Working Group ELDAH is dedicated to addressing the needs of the research & education community regarding the topics of legal issues and research ethics. In the age of digital information technology and the constant availability of information through the internet, it is not only important to have democratic access to knowledge, but also essential to consider issues of access, transparency, reusability, and attribution.

Of particular interest and concern to cultural heritage institutions and humanities scholars alike are issues of copyright on, provision of, and access to digitized source material. There is a recognizable political drive in the European Union to facilitate free and public access to cultural heritage and research data hosted at publicly funded institutions. However, the lack of legal harmonization and the diverse and often unclear national legislations on the use and provision of resources by public cultural heritage or research and education institutions has been prohibiting a much broader engagement. The recent EU regulation on Data Protection (GDPR) also caused a lot of insecurity and, due to lack of information and counseling, prompted frantic activity that often went beyond actual legal requirements and presents significant difficulties to conducting research. Therefore, ELDAH provides reliable information for humanities scholars facing legal questions and uncertainties in their research, not from a legalist but from a community- and practice-oriented perspective.

---

*Speaker

Beyond this legal scope, issues of ethical research practices and scholarly conduct are central to the Humanities and Social Sciences, perhaps even more so in a largely digital, internet-based research context.

Currently including about 40 members from 18 countries, covering a large variety of disciplinary fields, ELDAH produces recommendations (both to the DARIAH management team and the DH community at large), training and information materials on IPR, open licenses and Open Science in general, and offers workshops on these topics to scholars in the context of DARIAH events across Europe.

Additionally, ELDAH works in close collaboration with similar groups among the large European research infrastructures - like the Legal and Ethical Issues Committee of CLARIN ERIC (CLIC), the Consortium of European Social Science Data Archives (CESSDA) or the Europeana Networks Association - and pertinent working groups of national organisations - e.g. the Open Science Network Austria (OANA) and the Association for Digital Humanities in German-speaking Countries (DHd) - developing recommendations that will transcend national or infrastructural boundaries.

This poster will inform the audience about the main activities and topics covered by the ELDAH Working Group. It will give us an opportunity to address questions and concerns of the audience regarding legal and ethical challenges of their work directly, to share our recommendations, but also to collect valuable input on the current prevalent concerns of our community and to connect with working groups and initiatives on these topics beyond our current Eurocentric scope.

Keywords: Ethics, Privacy, Open Science, Licensing, Copyright

"Networking Women Translators in Spain (1868-1936) and their presence in Mnemosyne Digital Library"

Dolores Romero-López * [1], Jose González-Soriano [1], Lucía Cotarelo-Esteban [1], Alicia Reina-Navarro [2], María Martínez-Deyros [2]

[1] Universidad Complutense de Madrid – Spain
[2] Universidad Complutense de Madrid [Madrid] – Spain

Networking Women Translators in Spain (1868-1936) and their presence in *Mnemosyne* Digital Library

Studies of Spanish literature during the late 19th century and the first third of the 20th are evolving from research on canonical writers to the study of 'odd and forgotten' authors, themes and genres during what is now called *The Other Silver Age*. This research paper focuses on the work undertaken in the field of literary translation by the women writers of this period. After the 1868 revolution, there are a few timid attempts at mentalizing women on the necessity of scholarship to achieve the economic independence that will be the basis of their modern European mentality.

Based on the lack of critical studies published so far on gender and translation in Spain, this research firstly locates and lists the Spanish women translators and their translations. At a first stage these translators worked alone, in the solitude of their own rooms. During the first decades of the 20th century the new generation of women translators started to network by affiliation with the first women's associations, mainly the *Lyceum Club Femenino Español*, due to its international connections. Intellectuals and the general public could read the works of distinguished novelists, poets, playwrights and essay writers thanks to the translations carried out by these women translators. These women also showed their interest in the writings of contemporary women by translating their works into Spanish or glossing foreign ideas about how the modern woman should be, think or behave.

In their translations, valuable progress towards the awakening of a different female identity, committed to their social and cultural environment, was revealed. In addition, throughout the Silver Age period women dignified their work as translators by developing it professionally. From this perspective, translation became a sort of *affidamento* (Cigarini, 2000), an ethical attitude which serves to consolidate their feminine values.

It will be presented the Collection "Mujeres traductoras en la Edad de Plata", inside *Mnemosyne*. *Mnemosyne* is a Digital Library for Rare and Forgotten Literary Texts (*1868-1936)* whose purpose is to select, categorize and make visible in digital format literary texts that belong

---

*Speaker

to a forgotten repertoire in order to allow the historical review of the period. The collection of Spanish Women translators pretends to be a field of international experimentation for the creation of interoperable semantic networks through which a large group of scholars could generate innovative research and theoretical reading models for literary texts. See: http://repositorios.fdi.ucm.es/mnemosine/buscadorColeccion.php?name=Traductoras%20de%20la%20Edad%

*Mnemosyne*aims to be an open-access digital library allowing data modeling for **specific collections**(intellectual women, Madrid in the Silver Age literature, children kiosk literature, science fiction, photonovels, etc.) in support of research and teaching on Silver Age Spain. Through *Mnemosyne*it can be accessed digital edition of texts. The digitization of these works has been carried out by public and private institutions. The first version of the library is stored on the server of the Universidad Complutense de Madrid Library, itself linked to the collections of the digital library HathiTrust.40 In a search conducted in *HathiTrust*in 2012, the names, surnames, and pseudonyms of selected authors were used to locate a total of 2,873 digitized texts corresponding to "odd and forgotten" writers. The *Biblioteca Digital Hispánica*, which serves as the access portal for the digital collections of the *Biblioteca Nacional de España*, provided 2,448 works by male authors and 1,017 works by female writers. Now we are working on the **interactive editions**of some short stories with the support of the Spanish Biblioteca Nacional.

Behind the scenes of Mnemosyne's public presence online, the project is developing with the aid of the tool *Clavy*. *Clavy*is an RIA (Rich Internet Application) that is able to import, preserve, and edit information from collections of digital objects so as to build bridges between digital repositories and create collections of enriched digital content. *Clavy*also provides a basic system of data visualization, edition, and navigation. There are plans to integrate *@Note*, a collaborative annotation application, into *Clavy*. These two computational tools were developed by the software languages research group ILSA (http://ilsa.fdi.ucm.es/ilsa/ilsa.php) at the Universidad Complutense de Madrid.44 *Clavy*facilitates the import, export, and edition of records in multiple formats such as MARC2145, as well as their integration into Mnemosine's predesigned model with a view to their export into other compatible formats like XLS (Excel Binary File Format) or XML (Extensible Markup Language), or into other systems like OdA.46 Using *Clavy*, metadata for more than four thousand digitized objects from *HathiTrust*and the Biblioteca Digital Hispánica have already been imported into the *Mnemosyne*database. Logically, the data from these sources was described in MARC21, following the rules for library catalogs. The outcome was forseeable: in some cases, *Mnemosyne's*data model did not require the degree of detail furnished by MARC21, while in other cases it was necessary to incorporate new information absent from that format.

In conclusion, we wanted to demonstrate the richness of the translations produced by women in Spain, their commitment to modern women and the need to represent this approach through the Digital Library *Mnemosyne*in order to digitally reconstruct the Other Silver Age. Future works for the Clavy application will go through kernel evolution of the application to support distributed information through many datastores, extend data edition, enabling the use of external data edition tools and capability of importing the deep transformation of data operations scripts. Also we are improving a new cluster approach to increment browsing and finder protocols and services.

45

Mnemosyne, Digital Library for Rare and Forgotten Texts (1868-1936): -Collections and digital editions-

Dolores Romero * [1], Lucía Cotarelo-Esteban [1], Jose González-Soriano [1], María Martínez-Deyros [1], Alicia Reina-Navarro [2]

[1] Universidad Complutense de Madrid [Madrid] – Spain
[2] Universidad Complutense de Madrid – Spain

The objective of *Mnemosyne*, Digital Library for Rare and Forgotten Literary Texts (*1868-1936*) is to select, categorize, and make visible in digital format literary texts that belong to a forgotten repertoire in order to allow the historiografical review of the period. *Mnemosyne*has a repertoire of texts and authors who have remained in the shadow of the great literary figures of the first third of the 20th century. This digital library pretends to be a field of international experimentation for the creation of interoperable semantic networks through which a large group of scholars could generate innovative research and theoretical reading models for literary texts. See http://repositorios.fdi.ucm.es/mnemosine/queesmnemosine.php

*Mnemosyne*aims to be an open-access digital library allowing data modeling for **specific collections**(intellectual women, Madrid in the Silver Age literature, children kiosk literature, science fiction, photonovels, etc.) in support of research and teaching on Silver Age Spain. Through *Mnemosyne*it can be accessed digital edition of texts. The digitization of these works has been carried out by public and private institutions. The first version of the library is stored on the server of the Universidad Complutense de Madrid Library, itself linked to the collections of the digital library HathiTrust.40 In a search conducted in *HathiTrust*in 2012, the names, surnames, and pseudonyms of selected authors were used to locate a total of 2,873 digitized texts corresponding to "odd and forgotten" writers. The *Biblioteca Digital Hispánica*, which serves as the access portal for the digital collections of the *Biblioteca Nacional de España*, provided 2,448 works by male authors and 1,017 works by female writers. Now we are working on the **interactive editions**of some short stories with the support of the Spanish Biblioteca Nacional.

Behind the scenes of Mnemosyne's public presence online, the project is developing with the aid of the tool *Clavy*. *Clavy*is an RIA (Rich Internet Application) that is able to import, preserve, and edit information from collections of digital objects so as to build bridges between digital repositories and create collections of enriched digital content. *Clavy*also provides a basic system of data visualization, edition, and navigation. There are plans to integrate *@Note*, a collaborative annotation application, into *Clavy*. These two computational tools were developed by the software languages research group ILSA (http://ilsa.fdi.ucm.es/ilsa/ilsa.php) at the Universidad Complutense de Madrid.44 *Clavy*facilitates the import, export, and edition of records in multiple formats such as MARC2145, as well as their integration into Mnemosine's predesigned model with a view to their export into other compatible formats like XLS (Excel Binary File Format)

---

*Speaker

or XML (Extensible Markup Language), or into other systems like OdA.46 Using *Clavy*, metadata for more than four thousand digitized objects from *HathiTrust*and the Biblioteca Digital Hispánica have already been imported into the *Mnemosyne*database. Logically, the data from these sources was described in MARC21, following the rules for library catalogs. The outcome was forseeable: in some cases, *Mnemosyne's*data model did not require the degree of detail furnished by MARC21, while in other cases it was necessary to incorporate new information absent from that format.

Interactive Reading in Digital Collections

Dolores Romero [*] [1], Alicia Reina-Navarro [1], Aroa Rabadán [1]

[1] Universidad Complutense de Madrid [Madrid] – Spain

PROJECT ELITE-LOEP

The National Library of Spain (BNE) and the research group on Spanish Silver Age Literature (LOEP) at the Complutense University of Madrid present "La Edad de Plata interactiva" ("The interactive Spanish Silver Age"), a collaborative project to enrich the Library's digital collections and their use in teaching and research, exploring the resources that new digital technologies offer to the edition of texts in the field of cultural dissemination. The results of this research are inserted within the framework of eLITE-CM's Project (H2015/HUM-3426).

In response to new reading habits that the Digital Revolution has brought to the current panorama, the LOEP group of research has had as objective within the project eLITE-CM the development of three collections of forgotten texts of the Silver Age, rescued through digitization and enriched by computer programs, with the intention of exploring new resources that hypertext offers to philologycal studies in the Digital Age (Sanz and Romero, 2007).

To achieve our goals, it has been taken into account that dematerialization of cultural heritage linked to digitalization leads, necessarily, to new representations of cultural objects (Vinck, 2018: 73), which now overcome their physical limitations to become universally accessible through the Internet. This transformation, within the scope of text editing, implies a reconsideration of the book as a knowledge disseminator (Lucía Megías, 2012: 18) that now assumes characteristics such as transmediality or interactivity, which should be reinterpreted from an hermeneutic and phenomenological point of view, as part of a new perspective on the Reception theory linked to the field of Digital Humanities. Indeed, in terms of cultural dissemination, the enrichment of texts with images, sounds, concept maps, geolocators, hyperlinks, thematic transversality with other arts, etc. modifies the traditional concept of personal reading (understood as a unidirectional, silent and intimate situation), which now becomes interactive. If Bajtin (1975) considered that the literary work already constitutes a polyphony of voices, now the text enters into dialogue with the voices of the world that speak to us from the network.

OUR COLLECTIONS

Taking this into consideration, "La Edad de Plata interactiva" has carried out three collections of interactive books from the Silver Age linked to different topics, which will allow readers to access knowledge about this period through points of view not so widely considered by canonical historiography: 1) The Modern Woman in the Silver Age Literature; 2) Children's Literature in the Silver Age; 3) Madrid in the Silver Age Literature.

---

[*]Speaker

1) The "Modern Woman in the Silver Age" collection offers interactive editions of short stories, novels and essays written by modern Spanish women authors that, in most cases, were no longer read (except in the context of a very specialized criticism) and has been rescued and critically annotated for a general public thanks to technological resources. This collection is complemented with two monographs related to the topic of modern women, which allows users to access relevant information for the interpretation of the literary works and the historical period itself, carried out by specialists.

2) The "Children's Literature" collection presents two stories that has been updated for today's readers –most of them digital natives– in an interactive way. Both tales were originally published in Spain anonymously in 1923 by the publisher Calleja:*Plague of dragons*and *Spoiled Summer*. As a result of this research, our project discovered that the real author of these stories was the writer Edith Nesbit, one of the best English-language authors considered a background of current fantasy literature. In this collection we have implemented an audiobook system so that users can listen to the stories in a dramatized way. Readers will also be able to consult two magazines of additional information in which the historical context of the works have been studied and the evolution of dragon's myth in the fantastic literature has been traced (Reina, 2018).

3) In "Madrid in the Silver Age Literature" collection, we have selected several works from this period in which the city of Madrid presents itself as a space of modernity. For this purpose we propound a new approach to the works of the novelist and journalist Andrés Carranque de Ríos (1902-1936), who represented through his writings the complexities of a world in transformation towards modernity. This collection is complemented with a geolocator through which readers can access to several maps where the itineraries of the characters that appear in the texts have been traced, so that the different corners of Madrid during this historical period can be explored in a virtual way.

INSTITUTIONS AND FINANCING

This work has been funded by the National Library of Spain and by the project itself, within the Program of Research and Development Activities among research groups of the Community of Madrid in Social Sciences and Humanities, co-financed at 50% with the European Social Fund. The result of the research has resulted in an interactive kiosk of the Silver Age of Spanish Literature hosted on the website of the National Library of Spain.

LINKS OF INTEREST

- BNE interactive colections

http://www.bne.es/es/Colecciones/LibrosInteractivos/index.html

- Silver Age Literature kiosk:

http://cloud.madgazine.com/46f185c3185976675/?quiosco=46f185c3185976675&t=1542576418

- Press release**:**

http://www.bne.es/es/AreaPrensa/noticias2018/1118-edad-plata-interactiva.html

BIBLIOGRAPHY

BAJTIN, M. (1975). *Teoría y estética de la novela.* Madrid: Taurus.

HESS, E. D y LIDWIG, K. (2017). *Humility is the New Smart.* Oakland: Berrett-Koehler Publishers.

LUCÍA MEGÍAS, J.M. (2012). *Elogio del texto digital.* Madrid: Fórcola.

REINA NAVARRO, A. (2018). "Plaga de dragones reeditado con Madgazine". En *El universo mágico de Edith Nesbit.* Sevilla: Renaciamiento.

ROMERO LÓPEZ, D. (2014). *Los márgenes de la modernidad. Temas y creadores raros y olvidados en la Edad de Plata.* Madrid: Punto Rojo Libros.

SANZ, A and ROMERO LÓPEZ, D. (2007*). Literatures in the Digital Era. Theory and Praxis.* Newcastle: Cambridge Scholars Publishing

VINCK, D. (2018). *Humanidades Digitales. La cultura frente a las nuevas tecnologías.* Barcelona: Gedisa.

Keywords: Interactive Reading, Interactive Edition, Digital Collections

Project DESIR: How to Sustain an International Digital Infrastructure for the Arts and Humanities

Marco Raciti * [1]

[1] DARIAH-EU – DARIAH-EU – France

Europe has a long and rich tradition as a centre of research and teaching in the arts and humanities. However, the huge digital transformation that affects the arts and humanities research landscape all over the world requires that we set up sustainable research infrastructures, new and refined techniques, state-of-the-art methods and an expanded skills base. Responding to these challenges, the Digital Research Infrastructure for Arts and Humanities (DARIAH) was launched as a pan-European network and research infrastructure. After expansion and consolidation, which involved DARIAH's inclusion in the ESFRI roadmap, DARIAH became a European Research Infrastructure Consortium (ERIC) in 2014.

The Horizon 2020 funded project DESIR (DARIAH ERIC Sustainability Refined) sets out to strengthen the sustainability of DARIAH and help establish it as a reliable long-term partner within our communities. Sustaining existing digital expertise, tools, resources in Europe in the context of DESIR involves a goal-oriented set of measures in order to first, maintain, expand and develop DARIAH in its capacities as an organisation and technical research infrastructure; secondly, to engage its members further, as well as measure and increase their trust in DARIAH; thirdly, to expand the network in order to integrate new regions and communities.

The DESIR consortium is composed of core DARIAH members, representatives from potential new DARIAH members and external technical experts.

The sustainability of a research infrastructure is the capacity to remain operative, effective and competitive over its expected lifetime. In DESIR, this definition is translated into an evolving 6-dimensional process, divided into the following challenges:

Dissemination

Growth

Technology

Robustness

Trust

---

*Speaker

Education

With our poster, we would like to show how the project helps sustaining DARIAH.

Within DESIR, dissemination is the ability to communicate DARIAH's strategy and benefits effectively within the DARIAH community and in new areas, spreading out to new communities. Through the international workshops held at Stanford University and at the Library of Congress, DARIAH has been introduced to many non-European DH scholars. These events were an important first step to foster international cooperation between US and European colleagues as well as a catalyst for ongoing collaborations in the future. A third workshop will take place in Australia in March 2019.

DARIAH has currently 17 members from all over Europe. Nevertheless, efforts should be made to include as many countries as possible to bring in and scale, to a European level, even more state-of-the-art DH activities.

Six candidates ready for building strong national consortia have been identified, enabling a substantial expansion of DARIAH's country coverage. Additionally, thematic workshops are organised in each country as well as tailored training measures.

DESIR widens the research infrastructure in core areas which are vital for DARIAH's sustainability but are not yet covered by the existing set-up. As DARIAH expands across Europe, continuously enhancing and further developing the ERIC exceeds DARIAH's internal technological capacities. Two notable results were achieved so far: firstly, the publication of a technical reference as a result of a workshop organised in October 2017 with CESSDA and CLARIN. It's a collection of basic guidelines and references for development and maintenance of infrastructure services within DARIAH and beyond, addressing an ongoing issue for research infrastructures, namely software sustainability. Secondly, the organisation of a Code Sprint, focusing on bibliographical and citation metadata, which helped shaping DARIAH's profile in four technology areas (visualisation, text analytic services, entity-based search and scholarly content management). Another Code sprint is expected to take place in Summer 2019.

Another output is the implementation of a centralized helpdesk. This helpdesk is hosted by CLARIN-D and the solution of integration within the existing DARIAH website was the creation of a WordPress plugin. This plugin is used to connect our website with the OTRS server and allows the creation of issues easily by users unfamiliar with OTRS.

Sustaining a research infrastructure involves also two important aspects: trust and education. For DARIAH, it is crucial to increase trust and confidence from its users. In DESIR we develop recommendations and strategies accordingly, targeting new cross-disciplinary communities, based on the results of a survey and interviews addressed to the scientific community, with different levels of approach - national, institutional and individual.
In addition, education is a key area and the project contributes to the ongoing discussions about the role and modalities of training and education in the development, consolidation and sustainability of digital research infrastructures. We believe that investing time and efforts into training and educating users is a way of securing the social sustainability of a research infrastructure.

Keywords: Sustainability, research infrastructure, internationalisation, trust, training

The Standardization Survival Kit. For a wider use of standards within Arts and Humanities

Charles Riondet * [1], Dorian Seillier [1], Laurent Romary [1], Lionel Tadjou [1], Klaus Illmayer * [2],

Marie Puren [3]

[1] Inria de Paris – Institut National de Recherche en Informatique et en Automatique – France
[2] Austrian Academy of Sciences (OeAW) – Doktor-Ignaz-Seipel-Platz 2, 1010 Wien, Austria
[3] CNRS – UMR LARHRA – France

Arts and Humanities research has to address new challenges raised by the increasing amount of digital sources, contents and tools. New digital practices and protocols, new digital methodologies and services, new software and databases, offer a completely renewed framework for research, and encourage the emergence of a next generation of digitally-aware scholars.
European digital infrastructures, such as DARIAH and PARTHENOS, has thus acknowledged the growing importance to develop a data-centered strategy for the management of scientific data (European Commission, 2010). In this context has been developed the Standardization Survival Kit, or SSK (Romary et al., 2018, http://ssk.huma-num.fr) to help Arts and Humanities scholars understand the crucial role that proper data modelling and standards have to play in making digital contents sustainable, interoperable and reusable.

The Standardization Survival Kit is an overlay platform dedicated to promote a wider use of standards within Arts and Humanities. This comprehensive interface aims at providing documentation and resources concerning standards (especially authoritative references for each standard such as sources, Standard Development Organizations), and at covering three types of activities related to the deployment and use of standards in the Arts and Humanities scholarship: documenting existing standards by providing reference materials, supporting the adoption of standards, and communicating with all Arts and Humanities research communities.

It is a collection of research use cases (the " scenarios ") illustrating digital best practices in Digital Humanities and Heritage research, for both novice and advanced scholars in the domain of digital methods. It is designed as an easy-to-use online and collaborative platform with a user-friendly design. The idea of having general, end-to-end scenarios to help researchers carry out their project by following best practices and clear methods in their area of expertise is the most important design principle for the SSK, in order to effectively guide Arts and Humanities scholars through all available and relevant resources (collected within a dedicated Zotero library). The SSK scenarios come from reference best-practice use cases identified since the beginning of the project (PARTHENOS, 2016), as well as scenarios uploaded directly by researchers that

---

* Speaker

want to share their research good practices.

The SSK follows a "story-telling" approach to the use of standards in the Humanities and Social Sciences. The goal is to avoid providing yet another catalogue of resources, and to offer instead contextual, activity-based information on how to use standards for researchers who are unfamiliar with them, but could see how they are used and what workflow they help achieve by following a scenario.

The SSK is an open tool, imagined by and for the research community at large, but it aims at being integrated in the DARIAH ecosystem. It is now hosted by the National Coordinating Institution of DARIAH-FR: Huma-Num. Furthermore, its everyday intellectual maintenance, both on an institutional and content level, is better taken over by the DARIAH communities. **The DARIAH Working Group Guidelines and Standards**, chaired by Klaus Illmayer (OEAW, CLARIAH-AT), Marie Puren (CNRS, DARIAH-FR) and Charles Riondet (Inria, DARIAH-FR), aims at being a place where the SSK can be administered. It seems also very natural that the SSK becomes a place where methodological resources and best practices created or gathered by the various DARIAH Working Groups can be presented and archived. Doing so, we also pave the way for the SSK to be integrated into the **DARIAH Marketplace** and used by Arts, Humanities and Social Sciences researchers in the context of the European Open Science Cloud (EOSC).

The poster will show how an Arts and Humanities scholar can navigate the Standardization Survival Kit website, by taking the example of an actual reference scenario.

A live demo of the interface will also accompany the presentation, so that those interested in the poster will be able to search the website according to their needs. It will demonstrate how the SSK helps Arts and Humanities researchers select the most appropriate standards to perform a task or to carry out a research project successfully, but also how it helps a researcher or a team of researchers document their research good practices by creating and sharing a research scenario within the SSK. Instead of creating a simple catalog, standards are presented in context, within research projects in various fields of Arts, Humanities and Cultural Heritage research: History, Heritage Science, Linguistics, Literature, or Sociology.

More advanced users will also be able to edit the scenarios themselves, by submitting new resources or adding new steps. They can also create new scenarios. The SSK scenarios and steps can be easily extended, reused and customized, thanks to their flexible data model in TEI. A dedicated interface in the Standardization Survival Kit enables users to make suggestions, automatically converted in TEI according to the appropriate schema. The poster will present this interface and the associated functionalities. And for those who will be eager to test it, a live demo will be provided.

Bibliography

http://ssk.huma-num.fr

http://ssk.readthedocs.io

European Commission, Riding the wave : How Europe can gain from the rising tide of scientific data, report of the High Level Expert Group on Scientific Data, 2010, https://www.fosteropenscience.eu/content/wave-how-europe-can-gain-rising-tide-scientific-data

PARTHENOS, Report on user requirements, 2016, http://www.parthenos-project.eu/projects-deliverables/
Laurent Romary, Charles Riondet, Lionel Tadjou, Dorian Seillier. STANDARDIZATION SURVIVAL KIT (Final). [Technical Report] Inria Paris. 2018. hal-01925144

Barriers and Pathways to Engagement: Early Career Researchers and Research Infrastructures

Eliza Papaki * [1], Vicky Garnett * [2]

[1] DARIAH ERIC/Trinity College Dublin – Ireland
[2] Trinity College Dublin – Ireland

The European Commission notes that, in order to solve Europe's economic and societal challenges, innovation in science and technology, pursued through Research Infrastructures (RIs), is vital. Efficient RIs enable the greatest discoveries in science and technology, attract researchers from around the world, build bridges between research communities, allow the training of researchers and facilitate innovation and knowledge-sharing. As a pan-European network, DARIAH aims to enhance and support digitally-enabled research and teaching across the arts and humanities. And yet, despite this wide interdisciplinary and international reach, RIs such as DARIAH remain a distant concept to many of the researchers who could directly benefit from them.

The Community Engagement Working Group has, since Nov 2017, been investigating the often complex reasons for a lack of engagement with RIs among researchers as part of our wider research into Research Communities. Over the course of our funded project "Engaging Research Communities Beyond DARIAH", we conducted a webinar, an online survey, interviews with researchers at different stages of careers, and a roundtable session at a discipline-specific regional conference. Many of our respondents to these various methods of data gathering indicated a desire or willingness to participate in RIs such as DARIAH, yet barriers still exist that prevent them from doing so. One such area of interest included how these barriers may affect researchers by their career level, as well as disciplinary groups.

Researchers experience different pressures and issues at all stages in their career. However, in this proposal, we have decided to focus on Early Career Researchers (ECRs) as they are a target audience of DARIAH. The pressures on ECRs are well known, often taking on a sort of 'apprenticeship' role where not only the direction of the research is somewhat predetermined, but also membership of organisations and indeed RIs is decided by the more senior members in a team. An ECR does have some autonomy in terms of how they network, how they communicate with others in their field, and of course which teams or researchers they choose to work with in the first place.

We wanted to know in particular how communication methods may differ according to more nuanced community types. In doing so, we found differences in how ECRs communicate with

---

*Speaker

others in their field compared with mid-career and senior researchers, with ECRs favouring face-to-face communication over social media. When we looked at specific reasons for not engaging with RIs, the responses from the ECRs either indicated a lack of awareness about them, or focussed much more on time constraints due to competing priorities due to publication deadlines or anxiety over future employment.

Steps that RIs have taken to try to reach potential members have so far included creating national contact points to act as advocates, and by establishing some manner of training; either as occasional Summer or Winter Schools, or through online training resources. However information from and about these initiatives is often disseminated via social media, so despite these interventions there are still gaps between provision to enable engagement, and the modes of communicating these provisions.
This poster will present these results from the Community Engagement Working Group in more detail, and offer recommendations for how DARIAH members might integrate the needs of this specific research community into their wider communications practices.

Keywords: research communities, engagement, early career researchers, communications

Dispute over the data core in the humanities on the example of word formation database 'mem' in articles published in 'Teksty z Ulicy' in 2005–2018.

Robert Boroch [*] [1]

[1] University of Warsaw – Poland

The idea of 'humanities data' could be approached from different stands. Even those who treat this term, in humanities division, as a hypostatic or unrealistic object could provide important arguments supporting nonexistence such thing as 'humanities data'. That strong stand is based on a misunderstanding of the essence of humanities that differ from national academic tradition as well as social expectation from humanities itself. Social attitude towards humanities is linked with the way of life and ideology that determines the understanding of reality and urge fractions among social groups. That is why formulating an ontological as well as epistemological statements should deal with the explanatory power of 'humanities data' towards phenomena. Thus, this kind of approach is problematic that is bringing the problem according to the differences between an accurate knowledge about phenomena or an imaginative/interpretative knowledge on phenomena. Nevertheless, humanities products are recorded, and those records could be analysed in order to produce a methodological description of its change. In that, matters interesting issue is the material in which 'humanities data' are registered. Hence, we should talk here about the language and methods of propositional final judgment. According to this, I would present in my paper terms corpus analysis for Polish memetics in order to draw up the model concept of paradigm change in the scientific discipline on the example of terminology corpus used in the articles published in 'Teksty z Ulicy' between 2005–2018. I will divide my paper into two parts. In the first section, I will present the results of the quantitative analysis term based on lexical base -mem. In the second section, I will discuss the development of terminology corpus in order to predict possible changes in memetics paradigm represented in that journal. The major research hypothesis: **What is the saturation of the word base 'mem' in the articles published in 'Teksy z Ulicy' between 2005–2018?** Detailed research hypothesis: (1) **Is a growing level of saturation of the word base 'mem' means increasing tendency?** (2) **Is a steady level indicate its slowdown?** (3) **Is decreasing level of the terms based on 'mem' indicates dying idea**? I will support discussion with empirical material: chart 1 presents saturation with word formation database 'mem' in articles published in Teksty z Ulicy' in 2005–2018.

---

[*]Speaker

The user-friendly complexity: User experience design in the digital edition of The Correspondence of Skamander Literary Group

Paweł Rams * [1]

[1] The Institute of Literary Research of The Polish Academy of Sciences – Poland

1. Introduction – The Correspondence of Skamander Literary Group

In this paper I would like to present solutions used (by the New Panorama of Polish Literature team members) in digital edition of The Correspondence of Skamander Literary Group which were used during designing and re-organising structure of editorial informations and make it not only more intuitive, but also an estethic experience for user. This particular correspondence contains emigratory letters of two Polish XX-th century poets: Kazimierz Wierzyński and Jan Lechoń and their befriended publisher: Mieczyslaw Grydzewski. Project has been granted funding from gonvermental National Programme of Humanities' Development and two versions of edition are been preapering collaterally: a paper one and discussed in this poster: a digital one.

2. Home page – an invitation to digital exploring

Let's start with the most recognisable part of any website – home page, an opening for our projects. Simple, minimalistic google-style" main page with just a bare, white rectangle (aka the browser) as a starting point for your is users worth considering for linguistic, corpora/dictionary projects. When we prepare more complex collections like a corpora of letters, it isn't always clear to a new visitor of our websites what exactly he or she wants to search for specificlly, how the browser options work – and nothing is more irritating than a series or no matches found" messages.Thus, we have to divide the home page to a browser and exemplary documents. This provides quick access to the main content and encourages users to check the structure of the resources and documents. With the website's reload lists of those documents' reloads as well and next random ones appear. On Tei.nplp.pl, we included not only exemplary documents within a browser but also paths to collections of letters (we decided to divide the letter's corpora into smaller parts) and an index below exemplary documents section. In summary, it we do not necessarily need an opulence of features on the main page, however putting only a bare browser might be a fallacy as well.

3. An editorial commentary – how to divide (index) data

Digital edition of The Correspondence of Skamander Literary Group basis on earlier, professional printing scholarly edition. In the traditional, paper form all information and explanations

---

*Speaker

(so different as biograms, contextual footnotes or scholarly comments) have the same, linear narration form – footnotes. Our first step, and the same time challenge, was their dividing, functionalizing and – the most important – user friendly presentation. At the beginning we assigned 5 types of entities: persons, locations, works, journals and institutions/organisations. They are showed to the users in in two ways:
- As a list of grouped entities, which can be seen on the right side of any letter. Each entity types has their own colour, which is displayed in the content of a letter as well as on the list on the right. User can loosely tick one type, any, or none. However all entities previously marked by the editor are always active. Used solution allowed any user to read a letter in their own way (as a plain text or as fully or partially denoteted). Above and beyond thanks to colour identity user can easier find important information

- As a entities index. In this case user has possibility to search any kind of entity, their frequency in whole correspondence, use patterns, descriptions as well as varied derivations. Also here we stand up to intuitive and transparent way of viewing information.
Translation footnotes from printing to digital version was for us a big challenge. As I said before in the printing edition footnotes have the same visual representation regardless form the content. In digital edition we decided for three ways of presenting them, depending on their functions:
- Entity permanent description showed as a pop-ups always after clicking on the entity
- Contextual information, which have meaning only in one, specific context are showed above permanent description in a grey background
- Scholarly comments are presented as a traditional footnote (digit superscript). After clicking on the number jump the pop-up with content. In contrast to previous cases, user can not go from this level to the entity index

4. A chronological and logical order – a navigation through a corpora of letters

Even an avid reader may leave a website if its navigation isn't user-friendly. One of our methods to prevent this is to add breadcrunch, a fast shortcut in deep-structured collections on Tei.Nplp.pl. Althouh, it won't be enough to encourage users to stay with us after reading one letter (or article, etc.). That is why we created double navigation in Correspondence of Skamander's Poetry Group letters, specially designed for correspondence. The user can not only read those letters in the chronological order, but also in the referring-answer" order. It equips him or her with a straightforward path to the previous letter, to which the one currently read one is a reply.

5. A conclusion
To sum up. The Correspondence of Skamander's Poetry Group letters is a model example of traditional scholarly edition. But the same time we have to do with a really complex footnote structure, which have one line of 'narration' regardless of their diversity. Reader can not select interesting information in a simple way. Digital part of the staff wanted to translate this structure into more user friendly, and the same time easier form keeping the high value of editorial work. I am sure that our platform fuse the highest quality of digital edition with user friendly solutions. Furthermore we want give any user chance to make decision if he or she want to see quite simple structure or want to see any detail of The Correspondence. Our platform give that chance.

Keywords: user, centred design, letter, correspondence

Understanding and preserving the European Silk Heritage : how to produce accessible, interoperable and reusable Cultural Heritage data with the SILKNOW ontology

Marie Puren * [1], Pierre Vernus [1]

[1] LARHRA-CNRS – CNRS : UMR5190 – France

Silk was a major factor for progress in Europe, mostly along the Western Silk Road's network of production and market centres. Silk, however, has become a seriously endangered heritage. Although many European specialized museums are devoted to its preservation, they usually lack size and resources to establish networks or connections with other collections. The H2020 SILKNOW project aims to produce an intelligent computational system in order to improve our understanding of European silk heritage. This computational system is modelized and trained thanks to these datasets, mapped according to the SILKNOW ontology. In this paper, we will present how we have defined this data model, and how we have specified the entities to be represented by the ontology and the existing relationships between these entities. The design and implementation of the SILKNOW ontology representing the model is based on CIDOC-CRM.

Keywords: History, Digital humanities, museums, cultural artifacts digitisation, CIDOC, CRM, Cultural Heritage, textile, data modelization and mapping

---

*Speaker

The HumaReC project and its "web book"

Sara Schulthess * [1], Claire Clivaz [1], Martial Sankar [1]

[1] DH+, SIB, Lausanne – Switzerland

The aim of the poster is to present the HumaReC project and the concept and functionalities of its "web book". HumaReC is a project funded by the Swiss National Science Foundation that ran from 2016 to 2018 (https://humarec.org.). The HumaReC project was inspired by concrete experiences of the transformation in the publication process and the change of rhythm in Humanities research.
HumaReC has tested the change of rhythm by providing continuous publications of the data and research results. We have created a digital platform hosting a VRE that investigates the trilingual manuscript, Marciana Gr. 379. The codex contains the New Testament in Greek, Latin and Arabic and was created at the crossroads of Arabic, Byzantine and Norman cultures. Although this is the only New Testament manuscript known so far that presents the three languages together, the codex was never studied before. The VRE focused mostly on the visualisation of the manuscript and a continuous update of the transcript. In parallel, we regularly published research findings on a blog.

It became clear during the project that a different type of publication was necessary to communicate but also to validate our research. Long and structured writing belongs to the core skills of humanists. Yet, a conventional book would not have suited our research, notably because it would be impossible to connect the text with the refreshed data transcription continuously published on the VRE. To overcome this issue, we are developing a "web book" for HumaReC, in partnership with the academic publisher Brill. On the one hand, the reader of the web book (e.g. the scholar) will be able to link its content to the existing online material and data present on the VRE. On the other, the author (i.e. the researcher) will be able to write the text following the progression of his/her research in the VRE, while permitting the inclusion of readers' feedback. A first prototype of the web book about the Marciana Gr. 379 is under preparation. In the final phase, the web book will be peer-reviewed to ensure its quality, thereby affording it the same value as a traditional publication.

References

Clivaz, Claire, Sara Schulthess and Anastasia Chasapi, "HumaReC: Continuous Data Publishing in the Humanities". In: ERCIM News 111:21-22 (2017). https://ercim-news.ercim.eu/images/stories/EN111/E web.pdf.

Clivaz, Claire, Sara Schulthess and Martial Sankar, "Editing New Testament Arabic Manuscripts in a TEI-base: Fostering Close Reading in Digital Humanities. Journal of Data Mining & Digital Humanities". In: Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 1-6 (2017). https://jdmdh.episciences.org/paper/view?id=3700.

---

*Speaker

Eve, Martin Paul. Open Access and the Humanities. Contexts, Controversies and the Future. Cambridge: Cambridge University Press, 2014. https://www.cambridge.org/core/books/open-access-and-the-humanities/02BD7DB4A5172A864C432DBFD86E5FB4.

Fauchié, Antoine. "Le livre web, une autre forme du livre numérique". quaternum.net, 24 October 2016. https://www.quaternum.net/2016/10/24/le-livre-web-une-autre-forme-du-livre-numerique.
HumaReC – Humanities Research and Continuous Publishing. Research Platform of the Project. 2016-2019. https://humarec.org.

Towards an Ontology for European Poetry

M Luisa Díez Platas * [1], Helena Bermúdez Sabel [1], Salvador Ros Muñoz * [1]

Elena González-Blanco [2], Javier De La Rosa * [1]

Alvaro Pérez Pozo [1], Luciana Ayciriex [1]

[1] Universidad Nacional de Educación a Distancia – Spain
[2] Coverwallet – Spain

Introduction

An ontology is a formal and explicit specification of a shared conceptualization (Studer, Benjamins, & Fensel, 1998). From the point of view of ontological engineering, this formalization defines an abstract model that structures the knowledge of a specific domain (agreed upon by experts and interpretable by a machine) from its concepts and relationships. With ontologies, shared and distributed knowledge can be managed in such a way as toallow the integration of information from different data sets (Davies, Fensel, & Harmelen, 2003). An expression of this feature is the DBpedia, which contains multiple links to other data sets such as Freebase and Geonames, among others.

In this paper, we present the case of an ontology for the domain of European poetry. Given the fragmentary access to poetic resources (see González-Blanco & Selaf, 2014), an ontology, and more specifically an ontology network about European Poetry (EP), proposes a coherent and unambiguous semantic conceptualization of the domain. A shared conceptualization enables the integration of disparate poetic data sources belonging to different poetic traditions, overcoming well-known handicaps and idiosyncratic differences, such as the distinct organization of scholarship, or the heterogeneous bibliography.

Methodology

The starting point of the ontology construction was the analysis of different databases with contents related to one or more EP traditions.[1]As a result of this process, we built a data model comprising 41 entities, 490 attributes, and 407 relations.

In order to reduce some of the inherent complexity of the model, we decided to identify areas that could be organized within themselves. This allowed us to follow a modular design while ensuring higher cohesion and lower coupling of its modules. To this end, we extracted the ontological modules that represented each of these areas and that could form part of the ontologies of a network. Our goals were to make its structure comprehensible, to increase its capacity to be extended, and to allow other ontologies to selectively import modules from it. These modules are Work Module, Transmission Module, Structural Elements Module, Literary

---

*Speaker

Analysis Module, Prosodic Elements Module, Music Module and Additional Elements. All of them cover the different aspects referring to the poetic work, its representation, and the features to be analyzed. In addition, an autonomous upper ontology module (UOM) has been created that reuses other ontologies or modules of a similar nature, in which concepts and relationships independent of the domain that are defined, but which are necessary for the representation of complete knowledge such as information referring to Person, Event, Place y Role. In this sense, a module has also been defined to describe the treatment of dates in the ontology of POSTDATA.

For the definition of each module, an analysis of concepts and relationships has been carried out 1) to refine the semantics, 2) to solve design problems, and 3) to take the reutilization of existing ontologies into account.

Subsumption relations have been identified in order to refine the semantics and to help building hierarchies, such as prosodic information patterns.

The constructions identified to support use cases and structural relations are defined by means of design solutions such as ontological design patterns that allow to represent the mereological relations through the "part of" pattern,[2]or through the use of small ontologies such as "The Ordered List Ontology".[3]

The ontologies identified, in general, have not been completely reutilized. We have selected ontologies defined for analogous domains or related to the POSTDATA domain, or that cover semantic aspects of some of the modules. The ontologies and vocabularies used are, among others, Functional Requirements for Bibliographic Records (FRBRoo),[4]Schema.org vocabulary,[5]Europeana Data Model (edm),[6]CIDOC Conceptual Reference Model[7]. Then, the ontologies Friend of a Friend (foaf)[8]and Time[9]were selected for general purposes.

Much emphasis has been placed on the definition of restrictions that allow the establishment of requirements (mandatory or optional). Moreover, they express the cardinality of the object properties and data properties.

Conclusions and future work

A network of ontologies has been built which models the knowledge of European poetry. Each module covers a specific field of the domain. In addition, the use of ontologies and ontological design patterns enabled a good degree of alignment.

POSTDATA project aims at becoming a reference in terms of philological digital humanities standardization, interoperability by using linked open data(Gonzalez-Blanco et al., 2018).

Our proposal will support researchers in the organization of their information so that computational methods could be applied over their data. It will provide a means for the EP community of practice to easily exchange information, making their data shareable. In addition, it will enable the creation of applications that could retrieve, query, and manipulate these data.

This proposal is open to improvement and dialogue in the community and allows upgrades based on new knowledge derived from the great wealth and possibilities of the domain of poetry.

Acknowledgements

See Bermúdez et al. (2017) for more details about the resources analyzed and the methodology followed.

http://www.ontologydesignpatterns.org/cp/owl/partof.owl

http://smiy.sourceforge.net/olo/spec/orderedlistontology.html

http://www.cidoc-crm.org/frbroo/home-0

https://schema.org/

https://pro.europeana.eu/resources/standardization-tools/edm-documentation

http://www.cidoc-crm.org/

http://www.foaf-project.org/

https://www.w3.org/TR/owl-time/

Keywords: ontology network, Poetry, domain model, shared knowledge, linked open data

New Services and Tools for the Digital Monograph: the HIRMEOS Project

Andrea Bertino * [1]

[1] Göttingen State and University Library – Germany

New Services and Tools for the Digital Monograph: the HIRMEOS Project
Poster Presenter: Dr. Andrea Bertino (Göttingen State and University Library)

In the humanities and social sciences monographs are the researchers' laboratory and privileged means of communication. They enable scholars to deal in depth with differentiated and complex questions and help them to better define their academic profile. But although their impact and resonance can today be increased by digital tools and open access dissemination, monographs remain rather peripheral in the open science environment because of specific cultural, technical and economic barriers. The EU-project HIRMEOS aims to overcome these barriers by improving some digital platforms that already publish Open Access monographs. The guiding principle of the project is that making Open Access scholarly books more attractive for authors, readers and funding institutions means more than just releasing millions of Open Access documents in silos: it means creating bridges among countries and disciplines building an integrated trusted knowledge system.

HIRMEOS – High Integration of Research Monographs in the European Open Science infrastructure – is a 30-month project funded under the Horizon2020 Programme of the European Commission and is part of a wider infrastructure project, OPERAS, aimed at integrating Social Science and Humanities within the European Open Science Cloud and transforming collections of passive documents into rich and interlinked content. The project is based upon five Open Access books publishing platforms globally giving access to more than 10 000 books: OpenEdition Books, OAPEN Library, EKT Open Book Press, Ubiquity Press and Göttingen University Press. During the course of the project these platforms are being enriched with tools that enable

- identification, authentication and interoperability (DOI, ORCID, Fundref),

- entity recognition and disambiguation (by INRIA-DARIAH-EU entity-fishing),

- annotation of digital monographs (by Hypothes.is),

- certification of peer review processes (by DOAB),

- gathering of usage and alternative metric data.

---

*Speaker

The poster will give facts and figures about the involved platforms and their catalogs, the types of services implemented and the used standards, the expected benefit for the humanities and the social sciences and for the Open Science environment in general. More specifically, during our presentation we will focus on the results achieved in the work packages already concluded, i.e. those concerning the usage of **persistent identifiers** (WP2), the application of the **entity recognition and disambiguation service** in scholarly publishing (WP3), the **peer review certification system** (WP4), the implementation of a **tool for open annotations (WP5)**. To this end we will give information or practical demonstrations on

- how persistent identifiers improve the referencing of monographs and thus their discoverability and how the standardization of platforms by using the same persistent identifiers (ORCID, DOI, FundRef) enables better cross-linking between works and interoperability between different platforms;

- how the service of automatic entity mining (NERD) makes it possible to develop visualization services (wordclouds, facets, etc.) that have been tested on the platforms involved in the project and that can help the users to find Open Access works of their interest;

- how the specific peer review certification system used on the involved digital platforms can increase the confidence of authors and readers in the scientific quality of Open Access monographs;

- how the tool for open annotations can help to increase interaction with a monograph, build new communities, develop new didactic strategies, and also contribute to the realization of cooperative critical editions, a central task of digital humanities;

how the strategic cooperation between infrastructures for digital humanities such as DARIAH on the one hand and a distributed research infrastructure for scholarly communication such as OPERAS on the other can foster the impact and resonance of monographs in the humanities and social sciences

Keywords: Monograph, Open Access, Entity Recognition, Mass Annotation, Metrics, Peer review

Korpusomat - sharing open corpora

Łukasz Kobyliński * [1], Witold Kieraś [1], Maciej Ogrodniczuk [1], Zbigniew Gawłowicz [1], Michał Wasiluk [1]

[1] Instytut Podstaw Informatyki – Poland

Korpusomat is a platform that allows its users to create, automatically analyze and query text corpora for the purpose of linguistic research. The corpora may be created from various sources of textual data: texts uploaded by the user in one of many different file formats, downloaded from the Internet, or scraped from a particular website .The text is then analyzed automatically by the system, providing such annotation layers as morphosyntactic analysis and named entities. Furthermore, characteristic terms and keywords are extracted, as well as common word associations (e.g. adjective and substantive). Finally, the corpus may be queried by the user, using any of the provided annotation layers as query parameters.

Korpusomat has been successfully deployed as a stand-alone service (http://korpusomat.pl) to be used as a corpus creation and analysis tool, but also used to publish several existing text corpora, such as the Corpus of the 19. century Polish, KORBA (electronic corpus of 17th and 18th century Polish texts), or the one million word subcorpus of the National Corpus of Polish. The next version of Korpusomat (coming in mid-May 2019) will allow the user to share their corpora publicly or with specified users of the platform. This functionality will allow to publish open corpora, to be used by other researchers for comparison, or shared with a group of persons, for example class students.

Keywords: corpus linguistics, text annotation, morphology, named entity recognition, open data

---

*Speaker

New Panorama of Polish Literature (NPLP.PL) – How to tell digital scholarly stories

Bartłomiej Szleszynski * [1]

[1] Institute of Literary Research of the Polish Academy of Sciences – Poland

New Panorama of Polish Literature (NPLP.PL) is a platform for the presentation of research results in the digital environment in the form of scholarly digital collections. Each collection is telling a digital scholarly story" and using a different form to present content - it has its own narration, form, and visual identification. At this moment there are seven collections in Polish and two translated into English

New Panorama of Polish Literature (Nowa Panorama Literatury Polskiej) is also a team at Institute of Literary Research Polish Academy of Sciences one of the most important institutions in Polish humanities. Team's general goal can be described as translating traditional scholarly concepts into the digital environment – at the level of planning digital structures, creating software and finally entering content into those structures using the software that has been created. also responsible for the visual and typographical side of our collections. Unique, adequate to content form must be created for each of collections and under the full control of NPLP team. This is why interdisciplinary team of the New Panorama of Polish Literature includes not only literary and cultural studies researchers but also graphic designers and typographers. Furthermore NPLP since its beginning cooperates with the same team of programmers, who are not only very experienced in serious digital projects, but also in different forms of art.

The poster is going to show how NPLP platform and each digital collection is organized and how work of New Panorama of Polish Literature team is organized.

Keywords: digital scholarly stories, ogranizing work in DH, graphic design in DH

---

*Speaker

Data about DH training and education: the API for the DH Course Registry and its use cases

Hendrik Schmeer * [1,2], Tanja Wissik * [2]

[1] CLARIN ERIC – Netherlands
[2] Austrian Center for Digital Humanities (ÖAW) – Austria

One of the most important challenges facing DH today is [...] how to integrate and disseminate
knowledge, instead of merely capturing it and encapsulating it (cf. Tasovac et al. 2016). On
of the methods to make data and metadata available to different users and for different usage
scenarios is the implementation of Application Programming Interfaces. In the last decade the
development of APIs in Digital Humanities project have increased (cf. Tasovac et al. 2016) in
order to make the data and metadata more accessible.
Over the last years, the DH Course Registry, a platform to collect information on digital human-
ities programmes, courses, lectures and summer schools has grown and a respectable amount of
metadata has been collected (cf. Wissik 2019). Until recently, the collection of metadata was
only searchable and browsable via the online platform but the metadata itself was accessible
for analysis only to the DH Course Registry maintainer but not to external researchers or the
interested public. However, the possibility to access this rich source of data about DH training
and education would broaden the user base of the DH Course Registry.

The DH Course Registry received funding via the DARIAH Theme funding call 2018/2019 with
the project Digital Humanities Course Registry Sustain – Improving Sustainability through Us-
ability[1] until November 2019, to enable development towards greater sustainability. Ongoing
work now aims for the following goals: a better usability, user involvement and outreach, en-
hanced interoperability by researching the needs for and implementing an API. In the following
we will discuss the API development in more details and will present three sample use cases for
the API.

API Development

An existing, but previously undocumented API has been revised and made available already
last year[2].
So far the API has been implemented mainly to embed the DH Course Registry using an iframe

---

*Speaker

into foreign websites, filtering the output for a specific field, country or institution. It takes a country or institution as a filter parameter. The output is also available in JSON or XML format.

This API offers quite limited options and yet no API-driven data usage has ever been requested. Hence, we want to encourage the community to explore the hidden potential of the DH Course Registry also via the API. During the DH Course Registry Sustain Project, the API functionalities will be extended.

API Use Cases

As a starting point, we suggest three API use cases.

API use case 1
External usage of controlled vocabularies. We keep and maintain various category lists. Techniques and objects arose from and evolved from the TaDiRAH[3] taxonomy. The disciplines collection is a list of scholarly disciplines relevant to the field of digital humanities. The institution list provides a collection of universities with significance for digital humanities research and teaching.

API use case 2
Evaluating the usage of the tagging categories.
How various objects of research and disciplines relate to each other? How are teaching activities of specific fields distributed geographically?

API use case 3
How the teaching activities ingested into the DH Course Registry changed over time?

What would be your interest in our data, which can be made available by means of a more flexible API? Are there use cases beyond the original intent of the DH Course Registry?

During the poster and marketplace session, the team wants to discuss new proposals for API use cases and most wanted missing features with the audience and will collect the expressed interest in using and evaluating data available from the registry.

Furthermore, it will be possible, for those who do not have an account for the DH Course Registry yet, to register during the poster and marketplace session.

References

Toma Tasovac, Adrien Barbaresi, Thibault Clérice, Jennifer Edmond, Natalia Ermolaev, et al.. APIs in Digital Humanities: The Infrastructural Turn. Digital Humanities 2016, Jul 2016, Cracovie, Poland. pp.93-96, 2016, Digital Humanities 2016 Conference Abstracts.

Tanja Wissik, The role of research infrastructures in the area of DH education, training and competence building. 8th AIUCD Conference 2019 – 'Pedagogy, teaching, and research in the age of Digital Humanities', Udine, 23 – 25 gennaio 2019. pp. 233-235, 2019, AIUCD 2019 Book of Abstacts, http://aiucd2019.uniud.it/wp-content/uploads/2019/01/BoA-2019_PROVV.pdf

----

https://www.dariah.eu/2018/11/05/dariah-theme-funding-call-2018-2019-meet-the-winning-projects/

It is documented online here: https://registries.clarin-dariah.eu/courses/pages/downloads
http://tadirah.dariah.eu/vocab/index.php

Polish Literary Bibliography - New Research Data Portal for Complex Cultural Dataset

Beata Koper * [1], Tomasz Umerle * [1]

[1] The Institute of Literary Research of The Polish Academy of Sciences – Poland

Polish Literary Bibliography [PBL] is a large literary database containing app. 0,7 mln bibliographic records in a database form, and app. 1,5 mln records in a form of scanned pages of bibliography. It covers the period from 1944 to 2003 (database covers years 1989-2003).
It contains data on more than 40000 authors, 100000 works, and records close to 15000 cultural events, more than 10000 theatrical plays and radio plays and literary programmes, and 1000 films.

PBL has experienced two comprehensive remediations [Maryl and Wciślik, 2016; Bolter and Grusin, 2000] in recent decades. First, around 1998, it was transformed from a printed book into an online database (http://pbl.ibl.poznan.pl/) in open access. It was one of the earliest of such cultural databases in Poland, and it provided a stable environment for 20 years of continuous creation of rich bibliographic metadata. Yet, its Oracle-based production environment and its user interface were geared to faithfully represent the structure and layout of the bibliographic data of its printed predecessor, rather than to open for the possibilities of digitally-enabled data exploration, and as a result "it did not comply with any of the common standards in terms of record structure or data formats, what eventually lead to serious problems with both preservation and interoperability of collected data" [Maryl and Wciślik, 2006: 621].

From 2015 PBL has been transformed into a new kind of service (http://pbl.ibl.waw.pl), in cooperation between the Centre of Digital Humanities of the Institute of Literary Research of the Polish Academy of Sciences, the Institute's Department of Current Bibliography (creators of PBL), and IT specialists - Poznan Supercomputing and Networking Center. The redesign of the service was influenced by the research on users' practices via Google Analytics [Werla, 2016] and their practices (more than 700 responses to the online questionnaire), combined with the strong conviction that the efficient search engine of the service should be complemented by the rich data-exploration experience. As Whitelaw argues, it is the "rich, browsable interfaces that reveal the scale and complexity of digital heritage collections" [Whitelaw, 2015; Cf. Bomba, 2017]. The goal of the project team was to produce a service that can display not only the bibliographic records and the contents of authoritative files in the most seamless way possible, but also to open up the database for advanced research uses.

The poster will focus on the results of the second remediation of PBL -the redesign of the

---

*Speaker

bibliographic service, and the plans for its development. It will present 1. the main characteristics of the new PBL service and 2. the main challenges that the project team had to face and resolve that resulted from the complexities of the PBL dataset.

Design for rich data

In order to account for the complexities of PBL data, the new service includes:

> Many access points that enable exploration the complexities of the PBL dataset from different perspectives:
>
> search engine and advanced search engine,
>
> browsable finding aids that include a complex set of filters and facets; user can browse data starting with:
>
> subject headings,
>
> authoritative files: people, institutions, events, series, works,
>
> other categories: books, films, plays, TV programmes etc.

> Documentation describing PBL dataset in following manner:
>
> the scope of dataset,
>
> the organization of bibliographical work,
>
> the sources used for creating metadata.

> Provision for the use and reuse of the data in data-driven studies:
>
> API module (with rich documentation), and possibility of CSV downloads (in progress),

> Statistical module for data visualization (access via contact with PBL team):
>
> Kibana (https://www.elastic.co/products/kibana).

The main challenges for the project team

The main challenges that the poster will highlight are:

> finding the best workflow for collaboration between digital humanities researchers, IT specialists and bibliographers
>
> striking the balance between the need of consistent structuring of data and the

unstructured nature of annotations provided in specialised bibliography (insertion of tags with authoritative control into unstructured text was conceptualized and now is being implemented),

data processing for the data migration process:

introducing modern authoritative control and linked open data mark-up (e.g. combining 4(!) personal authority files into one, and restructuring institutional and event files), including the implementation of VIAF IDs (in progress),

parsing the annotated data fields in order to structure them (in progress).

retrodigitization of 40000 pages of bibliographic records (the goal was to scan and OCR them, and work conceptually on the possibilities of converting the data into the database format)

References

Bolter Jay D. and Grusin Richard (2000). Remediation: Understanding New Media. Cambridge: MIT Press.

Bomba Radoslaw (2017). Bazodanowe interfejsy. Projektowanie interakcji z dużymi zasobami danych kulturowych. Kultura Popularna 1(50).

Maryl Maciej and Wciślik Piotr (2016). Remediations of Polish Literary Bibliography: Towards a Lossless and Sustainable Retro-Conversion Model for Bibliographical Data. In: Digital humanities 2016: Conference Abstracts. Kraków: Jagiellonian University & Pedagogical University.

Werla Marcin (2016). Serwis internetowy "Polska Bibliografia Literacka". Analiza danych z Google Analytics za okres 7 miesiecy: od 1 kwietnia 2016 r. do 31 października 2016 r. Poznań. Online: http://lib.psnc.pl/dlibra/docmetadata?id=781&from=publication&showContent=true. Whitelaw Mitchell, Generous Interfaces for Digital Cultural Collections. Digital Humanities Quarterly 9 (1). Online: http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html.

Keywords: literary bibliography, cultural dataset, literary database, data visualization, user interface

The BackBone Thesaurus – evolving a meta-thesaurus for the humanities through collaborative, distributed thesauri maintenance and development

Eleni Tsoulouha * [1], Helen Goulis * [2]

George Bruseker [3]

[1] FORTH-ICS – Greece
[2] Academy of Athens – Greece
[3] Institute of Computer Science (ICS-FORTH) – N. Plastira 100, Vassilika Vouton, GR-700 13 Heraklion, Crete, Greece

The goal of this presentation is to showcase the development and collaborative management of the BackBone Thesaurus (BBT)[1] –an overarching thesaurus for the Humanities, under which specialist thesauri and structured vocabularies used across scholarly communities in the humanities can be aligned –along with the online tool that enables users navigate through the hierarchies of BBT, connect thesauri to it, as well as propose to modify the structure of the meta-thesaurus upon expanding its scope; namely the BBT Submission and Connection Management Tool (BBTalk).[2]
The BBT is the research outcome of work undertaken by the DARIAH Thesaurus Maintenance Working Group (TMWG); it is a faceted classification drawing on a small set of top-level concepts (facets and hierarchies) that are inductively obtained through a bottom-up analysis of vocabularies used in the fields of History, Archaeology, Ethnology, Philosophy of Sciences, Anthropology, Linguistics, Theatre Studies, Musicology and History of Art (Daskalaki & Charami 2017).

Taken together, the facets of BBT offer an open set of general classes that aim to broadly cover the domain of the humanities, while being clearly demarcated and completely disjoint from one another (TMWG 2015). The faceted classification technique is considered valid and reusable from a cross-disciplinary perspective. By providing an iterative and principled means to generate facets, this technique offers the major advantage of supporting open expansion of BBT into new areas of research in a sustainable and manageable fashion. Rather than limiting the basic set of terms, the basic terms are left in principle open to expansion and redefinition based on new input concepts.

In line with practices followed in thesaurus building, the concepts used in the BBT, as well as their relations, become salient through the definition of their scope notes –i.e. the properties singling them out from other concepts of the same level –plus the context they are found in –

---

*Speaker

i.e. their position in the overall structure (Doerr & Kalomoirakis 2000, Svenonius 2003). BBT only permits inclusion relations among its facets and hierarchies.

BBTalk is an online service developed to manage the function of integrating specialist thesauri with the meta-thesaurus, a process by means of which the top-level classification is corroborated or disconfirmed. If the latter, BBTalk allows users to propose changes to the BBT itself –submissions for changes involve the introduction of new high-level terms (corresponding to facets and hierarchies), plus the deletion, merging or splitting of existing ones.

These change submissions are then discussed among the various groups using BBT – its curators and specialist thesauri maintainers – until consensus is reached. In this context, the search for consensus revolves around the notion of user needs, defined by the relative match of local thesauri terminology clashes with the high-level classification proposed by the BBT. To support the process of consensus creation, a set of methodologically sound principles pertaining not only to thesaurus building, but also involved in the cognitive process of category building is invoked (Murphy 2002).

Furthermore, BBTalk allows users to browse through the facets and hierarchies of the meta-thesaurus and the list of terms from specialist thesauri that have been connected to it. Taken together, the functionalities of BBTalk allow it to work as a general maintenance system for the BBT (Georgis et al. 2019).

In the course of this presentation we will try to illustrate the processes and principles mentioned above by means of going through all the stages of thesaurus integration, in a stepwise fashion.

https://www.backbonethesaurus.eu/

www.backbonethesaurus.eu/BBTalk

REFERENCES

Daskalaki M. & Charami L. (2017). *A Back Bone Thesaurus for Digital Humanities,* ERCIM News 111, October 2017, Special theme: Digital Humanities , ERCIM News 111, October 2017 URL

Doerr M., & Kalomoirakis D. (2000). A Metastructure for Thesauri in Archaeology. Ljubljana, Slovenia, 18-21 April. BAR International Series 931, Archaeopress, Oxford, 117-126.

Georgis Ch., Bruseker G., & Tsouloucha E. (2019). *BBTalk: An Online Service for Collaborative and Transparent Thesaurus Curation,* ERCIM News 116, January 2018, Special theme: Transparency in Algorithmic Decision Making , ERCIM News 116, January 2018, URL

Murphy G. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Svenonius E. (2003). "Design of Controlled Vocabularies", in: *Encyclopedia of Library and Information Science*, DOI:10.1081/E-ELIS 120009038, New York, 822-838.

Thesaurus Maintenance Working Group, VCC3, DARIAH EU. (2015). *Thesaurus Maintenance Methodological Outline.* Greece. Retrieved from http://www.backbonethesaurus.eu/sites/default/files/working

GeoPolonistics: Digital Mapping of the Complexity of Polish Studies

Przemysław Górecki * [1], Mariola Wilczak * [1]

Olga Zakolska * [1]

[1] Instytut Badań Literackich Polskiej Akademii Nauk – Poland

- **Background**

This poster presents **GeoPolonistics**, a digital tool dedicated to mapping global Polish Studies in their complexity. It is related to "Polish Studies Newsletter" ("Biuletyn Polonistyczny": https://biuletynpolonistyczny.pl), a networking tool for humanities institutions, researchers, and students worldwide. Although the tool seems to be limited to a relatively narrow discipline of Polish Studies, we aim to present its unique institutional and collaborative considerations and prove it to be an example of a sustainable model for similar initiatives. The presentation demonstrates, with the use of ICT tools, a complex network of relationships in the humanities to a broader audience.

2. Key functionalities

The GeoPolonistics map is a visualization of relations between Polish and foreign institutions associated with Polish Studies. The primary idea of the tool is to integrate several forms of promotion. First of all, it serves as a database including information about Polish Studies and researchers, with references to a collection of articles and audio-visual materials to be found in "Polish Studies Newsletter". Secondly, it offers an interactive map of the world with an inscribed net of countries, cities, researchers and didactics. Forms of Polish Studies education, forms of educational and cultural activities conducted by a given institution and ways of cooperation with other Polish Studies centres are also listed in the database and indicated on the map.

The extended model allows each researcher, after authentication, to create and edit their individual profiles. It also enables an editor from a given institution to create and edit the institutional profile. Institutions in GeoPolonistics are expanded to a hierarchical level and related to educational offers and spatial data. In multi-faceted modules, this search engine for people, institutions and studies will be implemented as combined with an interactive map, which will also offer certain search capabilities (intuitive manipulation of the map). An interactive map with the above-mentioned content will be a port of the publicly available "Polish Studies Newsletter".

---

*Speaker

The whole interface will be designed by a graphic designer in line with current trends and functional requirements. It will be fully compatible with all devices (including mobile ones). Both the interface and content of the GeoPolonistics modules and the entire "Polish Studies Newsletter" will be available in Polish, English and Russian. Each module (Events, News, Articles and interviews, Scientific projects, Publications, Education and work, Institutions and Persons) existing in the "Polish Studies Newsletter" will be expanded with the possibility of defining individual fields in three languages. These contents, as well as the overall interface itself, will be displayed in accordance with the user's local settings.

Thanks to the implementation of diachronic mapping methods, consistently collected historical data will enable a visualization of the changes to which polonistics throughout the world has been subject over decades. A presentation of what Polish institutions around the world have in common aa well as how they differ will be our basis for research on the relationship between the development of Polish Studies and national cultures, and on the cultural history of Polish Studies with all its complexity.

As the project team, we are now at the conceptual phase, just before the implementation phase, the effects of which we will present during the DARIAH conference.

3. Institutional model

In this poster we would like to focus on the new cooperation model elaborated with the use of digital tools. Instead of a centralized editorial hierarchy, the newsletter taps into a dispersed collaboration model that is possible owing to the use of digital tools. Furthermore, it allows Polish Studies to gather a critical mass and solicit more interest as a research field which may also be understood as a contribution to the diversity of the humanities.

GeoPolonistics" calls for crossing lines and boundaries of the current humanities' reality, the digitizing nature of the world of science and of the changing nature of developing professional bonds. By crossing geographic borders, it embraces every research centre and reaches every interested researcher and recipient. Thanks to the remote cooperation model, the "GeoPolonistics" team also makes direct personal contacts with other research databases and teams, which results in a number of foreign queries and participation in international conferences. By addressing the needs of Polish Studies departments outside Poland we aim to foster interaction between the scholars and research varied approaches. The platform will have such features as organizing e-conferences and publishing their recordings on websites or creating conversation threads.

The system provides a communication platform for smaller, less renowned institutions, which can share their output globally, establish cooperation with interested partners and develop diversity in the scope of their research.

By means of international queries and conferences, the digital tool enables participation of researchers and students in Polish Studies abroad. Current experience of our team's foreign visits in Ukraine, Slovakia or the Czech Republic enables to raise the subject of the feedback of our project and its impact on academics, scholars and students. The groups we have cooperated with were interested in developing digital tools on their domestic grounds and they found the system on which Newsletter is based universal and practical in their specific local contexts of philological studies.

Towards a domain ontology dedicated to the description of anthropogenic traces in rock art

Alexandra Stoleru * [1]

[1] MAP-Aria – MAP UMR 3495 CNRS/MC – France

The Franco-Cantabrian region occupies an important place in the distribution of Paleolithic cave art in Europe. Successive discoveries have indeed led to extensive studies of cave art, involving teams composed of experts from different professional fields. The collaboration of these multidisciplinary teams implies manipulation of heterogeneous data sets. The inherent question concerns the semantic articulation of the produced elements and moreover the structuration of heterogeneous data.

How are the analytical elements produced by different experts semantically linked ? What are the terminology repositories and the controlled vocabularies of the domain ? How to structure and articulate this heterogeneous data in order to cross-reference it and thusly extract the potential knowledge encompassed ? Aiming towards a domain ontology, a knowledge conceptual model for the description of anthropogenic traces, this work states the adopted methodology.

Archeological documentation and controlled vocabulary represent effectively a consequential source of domain terminology, participating in the conceptual modelling of 5 top-level classes. Representing human-made markings, anthropogenic traces are at the heart of the proposed ontology. Their support, theme, representation technique and finally the alteration acting on the support represent the following classes. These are considered to be necessary and sufficient for the anthropogenic traces description. Their sub-classes are thereafter developed and questionned. These classes and their properties are to be merged with the CIDOC Conceptual Reference Model and its extensions.

The proposed ontology can serve as a basis for a possible information system and is a response to a need for semantic structuring of heterogeneous data. This articulation of the analytical elements produced within a domain is indeed interesting from the perspective of the extraction of new knowledge resulting from a crossing of information layers. This work is part of a larger project and is intended to be implemented within an information system. The conceptual modelling of the proposed classes remains nevertheless to be evaluated through experimentation by the experts in the field.

*Speaker

What Can The Research Data Alliance (RDA) Do For You?

Rene Van Horik * 1

1 Data Archiving and Networked Services – Netherlands

Abstract

The mission of the Research Data Alliance (RDA) is to build social and technical bridges to enable open sharing and re-use of data. The RDA vision is for researchers and innovators to openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.

The poster provides information on the activities and outcomes of the RDA that are potential relevant for the digitally enabled arts and humanities scholarship. The poster also aims to lower the threshold for scholars to participate in the activities of the RDA.

Poster Proposal:

The community-driven Research Data Alliance (RDA), launched in 2013, aims to build the social and technical infrastructure to enable open sharing of data. Over 8.000 people work together in working groups and interest groups to develop and adopt this infrastructure. Any individual or organization, regardless of profession or discipline, with an interest in reducing the barriers to data sharing and re-use can join the RDA.

The poster presentation consists of three parts. The first part concerns the provision of information on the background of the RDA and the way people can become active in the community. RDA members work together across the globe to tackle numerous infrastructure and data sharing challenges related to reproducibility, data preservation, best practices for domain repositories, legal interoperability, data citation, data type registries, metadata, etc. and these topics obviously are relevant for digitally enabled research in the arts and humanities.

Secondly, the poster provides an overview of RDA working groups, interest groups and their outcomes that might be relevant for digital humanities scholars. Working Groups are short-term (18 months) and come together to develop and implement data infrastructure, which could be tools, policy, practices and products that are adopted and used by projects, organizations, and communities. Embedded within these groups are individuals who will use the infrastructure and help in making it broadly available to the public. Interest groups are open-ended in terms of longevity. They focus on solving a specific data sharing problem and identifying what kind of infrastructure needs to be built. Interest groups can identify specific pieces of work and start up a working group to tackle those. Currently there are over 95 working groups and interest groups active in the RDA. Examples are "Digital Practices in History and Ethnography Interest

---

*Speaker

Group", "Linguistics Data Interest Group", "Empirical Humanities Metadata Working Group", "Education and Training on handling of research data", "Libraries for Research Data Interest Group" and the "Long Tail of Research Data Interest Group".

The third part of the poster concerns engagement with the arts and humanities community present at the DARIAH annual event to find out whether there is an incentive to set up a new interest group or working group on a topic that is not covered yet by an existing group. By providing an overview of the current activities and results of the RDA gaps might be detected.

The poster presentation is part of the "RDA Europe" project that acts as a plugin to the RDA and that provides funding for a broad range of grants, covering an early career program, grants for adoption for outcomes of the RDA and an ambassador program that facilitates this two-way communication and engagement between the digital humanities community and the RDA.

Czech Literary Bibliography: research infrastructure for Czech literature and literary science

Vojtěch Malínek * 1

1 Institute of Czech Literature, Czech Academy of Sciences – Czech Republic

The purpose of this poster is to present the datasets and the current research activities of the Czech Literary Bibliography research infrastructure (CLB; https://clb.ucl.cas.cz/), which is continuously operated at the Institute of Czech Literature of Czech Academy of Sciences since 1947. The CLB comprises a set of bibliographical and other specialized databases (literary figures, literary awards etc.) collecting the scientific informations on the Czech literature, literary culture and related disciplines. These datasets are updated on an ongoing basis and systematically reflect poetry, fiction, cultural journalism and specialist texts on Czech literature and literary life in the Czech lands from an ongoing printed production (dailies, periodicals, books) or newly also online sources (relevant electronic journals and websites etc.).

The parameters of the CLB's bibliographical databases (chronologically covering the period from 1770 to the present, dealing with approximately 1,500 titles and more than 2.2 million processed records) make them the most extensive specialist bibliography in the Czech Republic and one of the most complex sources of the literary-scientific informations in Europe. During 2014 CLB successfully underwent an assessment by the Ministry of Education, Youth and Sports of the Czech Republic and was included with effect from 2016 in the Roadmap of large Czech infrastructures for research, experimental development and innovation (this status was afterwards once more confirmed by the results of the evaluation in 2017).

Having used the internationally recognized standardized data format MARC21 and the set of persistent identifiers, the CLB data are available for the interconnection into the national and international networks for the exchange of the scientific informations.

This poster shall introduce the most important sources and activities of CLB, especially those related to digital humanities, such as:

The Retrospective Bibliography of Czech Literature and RETROBI system

This bibliography was originally processed as a card catalogue that chronologically covers the 1770–1945 period with more than 1.6 million bibliographical entries on Czech or other national literatures and other related disciplines. Retrospective bibliography is currently available online in the specialized system RETROBI (http://retrobi.ucl.cas.cz/).

This system for the digitization and online presentation of the card catalogues was developed

---

*Speaker

at CLB few years before and it interconnects the graphic informations (scanned cards) with the proprietary database structure, which can be further developed. RETROBI enables full text and semi-structured queries in the OCR representations of the original cards and includes a number of advanced user functions and utilities for administering and editing data (i. a. mass editing, editing of single database fields, user corrections of the OCR texts etc.). It has also been successfully deployed at other institutions and its further development is also a possible plan, particularly with regard to the potential conversion of Retrospective Bibliography data into a MARC 21-format structured database.

The "Contemporary" Bibliography

The database of contemporary" bibliography covers the period since 1945 up to the present and at the moment contains over the 600,000 entries. This dataset is based on historically or technically different collections of the data (e. g. particular parts of this database includes the Database of Czech Literary Exile or newly established databases of Czech Literary Samizdat and Czech Literary Internet), whose were few years before converted and unified under one roof of the internationally used librarian data format MARC21.

These two main CLB data sources are supplemented by **The database of Czech Literary Figures**, which is recording key biographical details of approx. 37,000 individuals involved in Czech literary life. Its data are systematically verified in the relevant sources (publicly available registries, archives etc.) or in the form of questionnaires.

This poster shall next to the main characteristics of CLB present its contemporary key projects, especially Czech Literary Internet and Bibliography of Czech Literary Samizdat.

Czech Literary Internet is an ongoing project realized in 2017–2021. Research activities undertaken as a part of this project contains i. a. following:

(a) bibliographical excerption from the Czech internet – the ongoing treatment of individual selected literary websites and portals, as well as information on the articles there and individual authors for the CLB databases. Every record shall include the link to the original web-page and – if possible – to its archived version.

(b) development of tools to analyse bibliographical data – a set of superstructural analytical and statistical tools shall be prepared to enable the end user to visualize the selected bibliographical data clearly (statistical graphs, analysis of the output of individual authors or websites classified according to the selected criteria, the modelling of these facts on a timeline etc.)

Bibliography of Czech Literary samizdat is connected with the intensive research of this phenomenon at the Institute of Czech Literature (particularly in December 2018 published Encyclopedia of Czech Literary Samizdat shall be mentioned). Treatment of samizdat sources have changed quite rapidly workflow of bibliographical process, because long time developed norms and schemas need to be adapted for this specific material.

Keywords: bibliography, databases, literary science, software tools

EURISE Network - Where DARIAH Meets Research Software Engineers

Yoann Moranville * [1], Tibor Kálmán * [2]

[1] DARIAH – DARIAH-EU – France
[2] GWDG – Germany

What is the problem with research software?
Sustainable management of research software and the application of software engineering principles are one of the biggest challenges that the operation of digital research infrastructures face in the Social Sciences and Humanities (SSH) community. Research software, an important pillar of such infrastructures, can be very heterogeneous, from simple source code to highly sophisticated software suites.

DARIAH being a pan-European research infrastructure also looks forward to bring prototype results to fully functional services and include those into its service landscape.

A possible solution: EURISE Network

To align technological recommendations and processes, the SSH infrastructures CESSDA, CLARIN and DARIAH have been working together. DESIR, an ongoing European Union's Horizon 2020 research and innovation project managed by DARIAH, made it possible to organise a joint workshop in 2017. Following this event, we have undertaken measures to align existing efforts towards a common understanding of technical requirements and recommendations. This includes the 'Software Maturity Modelling' developed by CESSDA, the 'Software Quality Guidelines' developed by CLARIAH-NL and the 'Technical Reference' originating from DARIAH's DESIR project.

The cooperation continues under the umbrella of the EURISE Network, where research infrastructures meet research software engineers, to strengthen the combined foundations for future collaborations. The main aims of the EURISE Network are: (1) collaboration on technology and development requirements & solutions for research infrastructures, (2) maintenance of a common set of guidelines, (3) teaching and education on matters of software quality, (4) support and engagement with the RSE community. With DESIR's help a second workshop could be organised in March 2019 jointly by the three ERICs. With about 40 attendees it was a great success and it helped to kick-off new collaborations.

How does the EURISE Network help DARIAH?

---

*Speaker

The Technical Reference (TR) of the EURISE Network was started by DESIR and has become the first common baseline of the three research infrastructures. It is a collection of best practices and guidelines for developers and maintainers of infrastructure components and can be used to evaluate the quality of ongoing developments or as a starting point for new research projects.

To make use of these guides, implementing an institutional manual is recommended as one will not find bounding choices in these documents but only recommendation whereas institutions or infrastructures might want to enforce some of those to reflect their use of certain technologies and tools.

For developers in the DARIAH ecosystem, it is suggested to follow these recommendations if possible. The guidelines are not set in stone, partner institutions can use these basic guidelines in order to create their own guidelines on a more specific manner used by their own software engineers.

How DARIAH does help the EURISE Network?

Until now, DARIAH has co-organised two EURISE workshops, taking advantage of a work package of its DESIR project, which provided the possibility to organise multi-ERICs workshops. EURISE proposes requirements and recommendations agreed between all three ERICs. DARIAH can help disseminate those to its members, partners and cooperating institutions.
Partner institutions in DARIAH's Virtual Competency Centre for e-Infrastructure (VCC1) contribute to EURISE and raise awareness of the benefits of the valuable outcomes of EURISE. VCC1 therefore draws on the expertise of other VCCs, for example, the VCC for Research and Education Liaison (VCC2), to ensure close collaboration with the Arts and Humanities Research community.

éLV - Digital Archives of Kasimir Twardowski: European Cultural Heritage, Enhanced Databases (TEI) and Philosophy of the Digital Humanities

Wioletta A. Miskiewicz [*] [1]

[1] CNRS/Archives Henri Poincaré Nancy – CNRS : UMR7117 – France

The importance of **éLV** Archives (acronym for the French name of Lvov-Warsaw School Archives) is first of all established on the exceptional character of the collection of manuscripts that was its starting point. Indeed, all those manuscripts from Kasimir Twardowski's office make us able to admire the last great collection, both difficult to access and unexploited, of the European philosophy at the end of 19th and in the first half of the 20th century. Twardowski's philosophical style is inspired by Franz Brentano, his master; it consists in critical analysis of philosophical problems by the conceptual criteria designed for this purpose, and aiming, if possible, at making them disappear. Once professor at the University of Lvov (at the time, this capital of the Western Ukraine was Polish),Twardowski put forward a philosophical agenda for *Mitteleuropa*, based on his profound conviction of the importance of philosophical training for a nation's culture. This is a program of scientific philosophy, where the analytic and rigorous method is privileged, as well as the criticism, the empiricism and the anti-irrationalism. During the thirty-five years of his academic life, Twardowski (1866-1938) educated several generations of Polish academic philosophers all assembled under the name of Lvov-Warsaw School; the most outstanding part of it is still the Polish School of Logic (Tarski, Lesniewski, Lukasiewicz, Jaskowski and others). Moreover, it is one of the sources not only of the tradition called *Austrian Philosophy* and of Karl Popper philosophy but of realist phenomenology of Roman ingarden et of *Denkstill* Theory of Ludwik Fleck. The reasons why such an important collection of writings remains unexplored until now is mainly political or geo-political (Miskiewicz, W.A. " Sur la philosophie polonaise au XXe siècle : autour d'un paradigme écarté ", *Revue des études slaves*,2014).

The **elv-akt** website http://www.elv-akt.net was created according to the terms of the convention signed on July 22nd 2005 in Warsaw by the Polish Academy of Sciences, the University of Paris (Paris 1) and the Center for Scientific Research (CNRS), concerning the "collaboration for the digitalization and the scientific operating and development of the archives of the Lvov-Warsaw School". The éLV website hosts the online digitalized manuscripts, developed the editorial activity related to these manuscripts and developed the scientific activity devoted to all aspects of the Twardowski school. The éLV editing and translation will provide mode XML/TEI.

It is the first French-Polish project in the field of digital arts and humanities. In 2014-2015, Wioletta A. Miskiewicz, the head of Project **éLV** actively worked for the preparation of the

---

[*]Speaker

candidature of Poland as a member of DARIAH.

Created at the IHPST (UMR 8590, Paris) the project moved (2017) to the Archives Poincaré in (UMR 7117, Nancy) and will undergo significant transformation: migration of contents on Omeka, creation of two blogs (*Pragmaties Vahier éLV* : https://elv.hypotheses.org/ and *Hyper-Twardowski Cahier métanumérique* : https://hyt.hypotheses.org/) and the development of a platform for crowdsourcing online transcription.

The experience of managing such polyvalent digital project, as well as the construction of enhanced databases (metadata, XML/TEI encoding, RDF etc) is a source of epistemolgical considerations on the domain of digital humanities.

With the gradual advancement of our digital area, the question of legitimacy and of fidelity to sources in digital humanities poses itself again more sharply. It is worsened by an incredible increase in the quantity of on-line archives and by the proliferation of fakes and hoaxes on the WEB. The ontological questions - in the both philosophical and IT senses - are inescapable.

What is an *object* in Digital Humanities?

There are platonic ontologies in DH proposed by Renear & Dubin and there are positivist ontologies: by Paul Caton (linguistic) or by Franco Moretti (cognitivist). Inspire by *actions and products theory* by Twardowski, we propose a semantic conception of the object in the digital humanities. Within the metaphysics of artefacts, we propose three categories of objects: digital document (DD), digital archives document (DAD) and digital research document (DRD).

Bibliography

W. Miskiewicz**,** in *Digital Humanities. Digital Ecosystems 3*, ed. W. Miskiewicz, N. Juchniewicz, DELab UW, Varsovie, 2015

W.A. Miskiewicz: " 'Text' *versus* 'Document'. Sur le platonisme dans les humanité numériques et sur la meïeutique TEI des textes", *The Linked TEI : Text encoding in the Web, TEI Conference and Members Meeting 2013*, UniversItalia, 2013, p.77-90.

Renear, A.H. & Dubin, D., "Three of the four FRBR group 1 entity types are roles, not types" in Grove, A. (ed), *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology (ASIST),* Milwaukee, WI., 2007.

Paul Caton " On the term 'text' in digital humanities", Literary and Linguistic Computing, vol 28, No.2, 2013, p. 209- 220

Franco Moretti *Graphs, maps, trees*, London: Verso, 2005
W.A. Miskiewicz in *Act-Based Conceptions of propositional Content*, Oxford UP, 2017

Facilitating the integration of DH methods into research on East European Cultures of Dissent

Piotr Wcislik * [1], Maciej Maryl * [2]

Jennifer Edmond [3,4], Wieneke Lars [5], Labov Jessie [6], Pim Van Bree [7], Geert Kessels [7]

[1] Institute of Literary Research, Polish Academy of Sciences – Poland
[2] Institute of Literary Research, Polish Academy of Sciences (IBL PAN) – Nowy Swiat 72 00-330 Warsaw, Poland
[3] Trinity College Dublin – Ireland
[4] Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France
[5] University of Luxembourg – Luxembourg
[6] Central European University – Hungary
[7] Lab1100 – Netherlands

With this poster, the international scholarly network *New Exploratory Phase in Research on East European Cultures of Dissent* (NEP4DISSENT) wishes to invite collaboration in **facilitating the integration of DH methods and tools by the multidisciplinary community built around the study and curatorship of the cultural legacy of resistance and dissent in former socialist countries in comparative and transnational perspective**. Resistance and dissent in former socialist Europe 1945-1989 constitutes a remarkable chapter of Europe's recent past, which not only informs in a decisive way the identities of post-socialist societies, but has also reshaped the continent as a whole and still provides an important reference for contemporary social movements worldwide. NEP4DISSENT (COST Action 16213, 2017-2021) is an EU-funded international scholarly network formed by around 200 participants from 38 countries to date. **The main aim of NEP4DISSENT is to trigger the next discovery phase of this legacy by forging a new, reflexive approach, and by providing a platform for incubating networked, transnational, multidisciplinary and technology-conscious research with creative dissemination capacities.** NEP4DISSENT creates a valuable interface for communication between three communities of practice. Facilitated by IT experts with humanities and social sciences expertise, the network enables participant researchers to train with cutting edge digital tools, and to increase their capacities for creative dissemination through engaging in productive dialogue with art and cultural heritage curators, in order for future research to be technologically advanced and better disseminated (please find out more at nep4dissent.eu).

The research and capacity building agenda of NEP4DISSENT represents a **complex and original challenge for the marketplace of digital research infrastructures due to its multidisciplinary character, the uneven propagation of DH research practices** between

---

*Speaker

disciplines and national scholarly communities East and West, **the uneven digital readiness of the often very ephemeral sources, as well as its multilinguality** that makes the comparative and transnational perspectives difficult to apply. On the other hand DH approaches, in particular methods of data aggregation and federation, techniques of text, image and layout recognition, as well as tools for network-graph and spatio-temporal visualisation and analysis, are u**niquely qualified to explore in full scope that comparative and transnational dimension** of the dissident networks of solidarity, which has been one of the most extraordinary aspects of that legacy.

NEP4DISSENT Working Group 5 *Mediating Research through Technology*, involving representatives of projects such as CENDARI, COURAGE, NODEGOAT, IMPRESSO, as well as research infrastructures: DARIAH (DiMPO Working Group), or CLARIN, has been set up to tackle the complexity of this scenario. The mission of WG5 is to map the specific needs of the network's participants against the available digital research infrastructures, to facilitate DH capacity building through training and research collaboration, to offer recommendations for new or improved services that would better respond to the specific problems of infusing the digital at various stages of the life-cycle of scholarly and curatorial projects in this domain, and most broadly, to promote and deepen our understanding of the role of the digital technologies in research on contemporary history.

The poster **will feature the conclusions drawn from the exploratory activities of WG5 conducted thus far**, characterised by the **multi-channelled, ethnography-inspired approach to understanding the place of technology in the existing research workflows** of the NEP4DISSENT participants, and based on a variety of inputs, including two surveys circulated in December 2017 and summer 2018 among the network participants supplying a broad baseline understanding of the network and the concerns of its contributors; structured group interviews conducted in February-May 2018 with key network members, whose profiles and interests indicated that they occupy research positions at the interface between analogue and digital sources and approaches; and informal conversations during the NEP4DISSENT assemblies in Brussels, Warsaw and Belgrade.

Further, the poster will **stimulate discussion about training schools as a fulcrum of DH methods and tools propagation**, welcoming insight about best practices which will help to better prepare for the course *Cultures of Dissent in Eastern Europe (1945-1989): Research Approaches in the Digital Humanities* organized by NEP4DISSENT in the framework of a Central European University Summer School immediately after the DH2019 conference.
Finally, the poster will create a **meeting point for both digital research infrastructure service providers** who would like to expand their user base towards this intriguing domain of contemporary history scholarship, and for **experts interested in broader methodological, technical and social aspects of penetration of DH approaches** in both similar and different scholarly communities of practice.

Keywords: Dissent, Eastern Europe, digital research workflows, knowledge transfer

Changing historiography into a database: proposition of the project Prosopographia Polonorum.

Adam Zapala * [1]

[1] Institute of History, Polish Academy of Sciences – Poland

A crucial issue of historiography is to describe and analyse human lifepaths. For more than hundred years, historians were fulfilling this task by publishing their works in printed form (e.g. monographies, source editions, biographies, etc). Nowadays we have plenteous and diverse corpus of historiographical texts, containing great knowledge, but impossible to search for historians who would like to carry out quantitative or broad-scale comparative research.

The main aim of the Prosopographia Polonorum project is to collect all the available information on the careers of the late medieval Polish in an easily searchable database. Nowadays there are dozens of prosopographical monographies, but every comparative or quantitative research requires that each researcher develops his/her own database. This situation is a great obstacle for Polish historiography. Prosopographia Polonorum would imply the creation of a complementary database containing lists of people with their social circumstances and family origin, the offices they held and the property they had in a specific time. This database would be based on a corpus of texts scanned, optically recognized (OCR) and tagged in XML.

The initial plan assumes to use a corpus of historical studies concentrated on careers (mostly biographies and prosopographical lists of officers). However, the corpus could be expanded with time to printed source editions and even original scanned documents. The database could also include unpublished notes of historians. In the final form, the database would be transformed into an open multimedia platform for research on the late medieval Polish society, also benefiting richly from crowdworking.

Keywords: History, Late Medieval, Prosopography

---

*Speaker

PARTHENOS Guidelines to FAIRify data management and make data reusable

Francesca Morselli * [1,2]

[1] Data Archiving and Networked Services – Netherlands
[2] Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

For the past couple of months, a team of PARTHENOS project members has been working in close collaboration on the creation of the *Guidelines to FAIRify data management and make data reusable.*

A comprehensive set of **PARTHENOS Guidelines to FAIRify data management and make data reusable** is focusing on the topic of common policies. This compact guide offers twenty guidelines to align the efforts of data producers, data archivists and data users in humanities and social sciences to make research data as reusable as possible based upon the FAIR Principles. The guidelines are grouped around the letters of the F-A-I-R principles and cover topics such as data preservation, metadata completeness, copyrights and reuse. Each guideline has recommendations for both researchers and archives as it is recognized that different priorities may apply to each case.

The guidelines result from the work of over fifty PARTHENOS project members. They were responsible for investigating commonalities in the implementation of policies and strategies for research data management and used results from desk research, questionnaires and interviews with selected experts to gather around one hundred current data management policies (including guides for preferred formats, data review policies and best practices, both formal as well as tacit)

The work was coordinated by Hella Hollander, Femmy Admiraal and Francesca Morselli. The graphic design of the PARTHENOS Guidelines was developed by Verbeeldingskr8 and is available under a CC-BY 4.0 license here.

Keywords: PARTHENOS, guidelines, FAIR, data management, reuse

---

*Speaker

Openness and access to data

Archaeology and open data: a step towards decolonization

Marta Lorenzon [*] [1]

[1] The University of Helsinki (Centre of Excellence ANEE) – Finland

Archaeological data are not a single unified entity as archaeological reports include a wide variety of different types of primary data, usually recovered during excavation (i.e. quantitative and qualitative). The reality of archaeological data is complex and multifaceted: firstly, archaeology is a multidisciplinary endeavor and as such it includes a variety of different data types (i.e. geochemical, osteological, spatial, climatic data). Secondly, the increase use of digital technologies in excavation helps to create primary data that make the archaeological records progressively more accurate, and theoretically accessible (Kansa and Whitcher-Kansa 2013). Unfortunately, this is not always the case as primary data are often not disseminated through reports or scholarly publications (Kansa and Whitcher-Kansa 2013;Kintight et al. 2014; Marchetti et al. 2018). While archaeology *per se* is the collection of fragmentary information from material culture, which often entails putting together incomplete records to interpret past societies, the main goal of archaeologists is to try, and answer research questions based on this information. Thus, any type of primary data becomes increasingly important to understand and solve, at least partially, this complex puzzle. This latter aspect makes the sharing of primary data of utmost importance within the scholarly environment, but in most cases what it is shared in publications and/or reports is just the interpretation of the evidence and not the raw data by itself (Kintight et al. 2014; Marchetti et al. 2018). In the past, this may have been partially due to the difficulties in creating repositories for multidisciplinary primary data or presenting them in one singular platform, but digital technologies have addressed partially these issues (Kansa 2012; Gidding et al. 2013; Lorenzonand Nelson-Vijoen 2016). Simply put, as archaeological research questions are answered by a combination of records based on the collaboration of different specialists, interdisciplinary repositories that can guarantee long-term preservation as well as immediate publishing are a positive asset for the discipline that should be promote and encourage as they sponsor both open data and open access (Lorenzon and Nelson-Vijoen 2016).

A secondary issue that is part of decision-making process of data publishing regards ownership of data. Does the data belong to the archaeologists? Does it belong to the country where the excavation is located? Or to the local community nearby the excavation location? Or the financial sponsor of the excavation? As in Europe most national funding bodies as well as European grants are financed by public money, that should guarantee open access to the primary data as part of public interest and transparency policies (Marchetti et al. 2018). Legally though the questions regarding ownership of data can have multiple answers depending on the legislation of different countries. Morally and ethnically, I argue archaeological data should belong to everyone as a source of public information about our shared past. This is why open data, and therefore open access publishing, are key instruments to achieve a real open science. Furthermore in

---

[*]Speaker

the current political climate, where academics are finally realizing the importance of decolonizing archaeology, open access policies also allow wider access to these resources by scholars from developing countries as well as local and indigenous communities with no additional cost. These ethnical questions are pivotal points in the process of analyzing and publishing data from archaeological research and can impact their long-term preservation (Kansa, Whitcher-Kansa and Watrall 2011; Lorenzon and Nelson-Vijoen 2016). Data preservation (i.e. storage security, durability and longevity) is of essential significance not only to promote open science, but to guarantee accessibility for future generations of scholars.

This paper rises questions regarding the relationship between archaeological practices and publication of primary data, focusing on the ways open data can be a key determinant in the decolonization of archaeology and presenting possible solutions to address and determine policy practices in developing an interoperable system for archaeological, and more generally humanity research.

References

Gidding, A., Levy, T. E., and DeFanti, T. A. (2014). ArchaeoSTOR: The development and utilization of a web-based database for the field and lab. *Near Eastern Archaeology, 77*(3), 198-202.

Kansa, E. C. (2012). Openness and archaeology's information ecosystem. *World Archaeology* 44: 498–520.

Kansa, E. C., and Whitcher-Kansa, S. (2013). We all know that a 14 is a sheep: Data publication and professionalism in archaeological communication. *Journal of Eastern Mediterranean Archaeology & Heritage Studies*1: 88–97.

Kansa, E. C., Whitcher-Kansa, S., and Watrall, E. (eds.) (2011). Archaeology 2.0: New Approaches to Communication and Collaboration, Cotsen Digital Archaeology Series 1, Cotsen Institute of Archaeology Press, Los Angeles. Available at: http://escholarship.org/uc/item/1r6137tb.

Kintigh, K. W., Altschul, J. H., Kinzig, A. P., Limp, W. F., Michener, W. K., Sabloff, J. A., Hackett, E. J., Kohler, T. A. Ludäscher, B., and Lynch, C. A. (2015). Cultural dynamics, deep time, and data: Planning cyberinfrastructure investments for archaeology. *Advances in Archaeological Practice* 3(1): 1–15.

Lorenzon, M. and C. Nelson-Vijoen. (2016). "Interdisciplinary methods of data recording, management and preservation", In S. Campana, R. Scopigno, G. Capenitiero and M. Cirillo (eds) *CAA2015 "Keep the revolution going" Proceedings of the 43rd Computer Applications and Quantitative Methods in Archaeology:* 187-90.Archaeopress: Oxford
Marchetti, N., Angelini, I., Artioli, G. et al.(2018). NEARCHOS. Networked Archaeological Open Science: Advances in Archaeology Through Field Analytics and Scientific Community Sharing". In *Journal Archaeolical Research* 26 (4): 447-469.

Keywords: decolonization, open data, archaeology, open science

OpenAIRE services for Open Science

Alessia Bardi * [1], Miriam Baglioni * [1]

Argiro Kokogiannaki [2], Pedro Principe [3], André Vieira [3], Paolo Manghi [4]
Erzsébet Tóth-Czifra [5]

[1] Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Italy
[2] ATHENA - Research and Innovation Center in Information, Communication and Knowledge
Technologies – Greece
[3] University of Minho [Braga] – Portugal
[4] National Research Centre – Italy
[5] DARIAH – Germany

According to Open Science (OS) principles, all scientific products generated by research activities (e.g. scientific literature, research data, software, experiments) should be made available as soon as possible "under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods".[1]
The effective implementation of Open Science is today hindered by several cultural and technical barriers. Publishing and reward policies are still focused only on research literature (books, journal articles, etc.). Attention to research data is growing in some communities of practice, while the publishing of research packages representing digital experiments is still not a usual procedure. The digital tools and services provided by Research Infrastructures and used by researchers are not connected to the scholarly communication ecosystem: publishing is therefore still a manual action, typically performed by researchers at the end of a research activity. Research products are made available on several scholarly communication data sources (e.g. repositories, archives, which may be institutional repositories, community-specific archives, generic repositories), making it hard to keep up-to-date links between products (e.g. research data reused by several studies) and to identify those that are relevant for a specific research community or that have been produced using services and tools provided by a given research infrastructures.

The OpenAIRE initiative (www.openaire.eu) is a legal entity representing a partnership of more than 50 institutions working to promote and support a sustainable implementation of Open Access and Open Science policies for reproducible science, transparent assessment and omni-comprehensive evaluation.

It supports the implementation and alignment of Open Science policies at the international level by developing and promoting the adoption of global open standards and interoperability guidelines[2]to realize a sustainable, participatory, trusted, scholarly communication ecosystem, open to all relevant stakeholders (e.g. research communities, funders, project coordinators) and capable of engaging society and foster innovation.

Thanks to the network of National Open Access Desks (NOADs), OpenAIRE supports the

---

*Speaker

99

implementation of Open Science at the local and national level, supporting researchers, project coordinators, funders and policy makers with training and support activities like workshops and webinars.[3]

Furthermore, OpenAIRE features a technical infrastructure that materializes an open, deduplicated, participatory metadata research graph of interlinked scientific products (including research literature, datasets, software, and other types of research products like workflows, protocols and methods), with access rights information, linked to funding information, research communities and infrastructures. The graph is materialized by collecting more than 100 millions of metadata records from more than 9,000 scholarly data sources world-wide.[4]In addition to the information collected from scholarly data sources, the graph includes metadata information and links between products that are *(i)*asserted (i.e. claimed) by users of the OpenAIRE Explore portal via the "Linking" functionality, and *(ii)*inferred by full-text and metadata mining algorithms. In particular, OpenAIRE algorithms enrich the graph with:

- Links from research products (literature, data, software or products of other types) to project fundings;

- Links from research products to research infrastructures;

- Links between research products (e.g. from literature to research software and data);

- Subjects and keywords based on standard classification schemes;

- Authors' affiliations;

- Countries for which the product is relevant;

- Research communities for which the product is relevant;

- Citations.

The resulting graph is called the *OpenAIRE Research Graph*and it is openly available under CC-BY license[5]to programmatic clients via the Develop API:https://develop.openaire.eu. Added-value services are also built on top of the graph to offer Open Science services to different stakeholders of the scholarly communication ecosystem.

Researchers can deposit any kind of research product on Zenodo (www.zenodo.org), the OpenAIRE "catch-all" repository, exploit the OpenAIRE Explore portal (https://explore.openaire.eu) to discover research products and check if there is an Open Access version available. Moreover, the Explore portal offers a reporting tool to *project coordinators*, who can also view statistics about the research products funded by a given project.

Content providers, like institutional repositories, have access to a plethora of services that support them in the implementation of interoperability guidelines, in the collection of usage statistics and for metadata enrichment via the Content Provider Dashboard (https://provide.openaire.eu).

Research communities can benefit from a dedicated gateway where researchers can find all the products that are relevant to a given community and have easy access to Open Science publishing tools thanks to the Research Community Dashboard (https://connect.openaire.eu).

Funders and *research infrastructures* can monitor their research impact, the compliance to their Open Access mandates, and the uptake of Open Science publishing practices among the researchers they funded via the OpenAIRE Monitor portal (https://monitor.openaire.eu).

FOSTER definition of Open Science: https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

OpenAIRE Guidelines for Content Providers: https://guidelines.openaire.eu

OpenAIRE support and training on Open Science: https://www.openaire.eu/support

In the OpenAIRE beta infrastructure -https://beta.explore.openaire.eu. The production infrastructure (https://explore.openaire.eu) collects more than 30 millions of metadata records from more than 13,000 scholarly data sources.

Some data sources aggregated by OpenAIRE do not allow to re-distribute metadata in CC0; the availability of another version of the graph including only information that can be redistributed in CC0 is in OpenAIRE's plans.

Access and Use of Smart Hospitals Data for Research Purposes

Jaroslaw Greser * [1]

[1] Adam Mickiewicz University in Poznań – Poland

The Internet of Things (IoT) is a revolution happening around us. Devices, system components and networks are becoming autonomous, ubiquitous and interconnected. It is estimated that by 2020 the total number of IoT units reach 25 to 30 billion while the amount of data they are going to generate will reach 500 zettabytes. They are widely used both in services and industry and the health system is considered to be the most interested in aiming to deliver optimal patient care. This is particularly evident in hospitals which tending to become 'smart hospitals'. That means IoT components are supporting the core functions of a hospital.

The IoT medical devices collect, process and store different types of information to ensure efficient and effective surgical and diagnosis processes with low error rate and cost-effectively. To achieve it they accumulate medical data and personal information. Nevertheless, at the same time, they gather records of regarding life habits, financial resources and sensitive personal information not necessary for medical treatment like political views or sexual orientation. This applies to patients, relatives and hospitals' staff.

This information may prove useful in public decision-making in such areas as public health, law enforcement, public safety or social policy. They are also crucial for evidence-based research in humanities not only related health. Nevertheless, those data are either inaccessible to the scholars or access is limited. At the same time, data is collected by the manufacturer of the device, both to enable analyses of its functioning as well as to develop consumer profiles or predict particular behaviours of the users. Data thus amassed constitutes an asset which possesses economic value and as such, it is protected by proprietary rights.

In my paper, I would like to address two issues. First, I am going to discuss possibilities of accessing data based on mechanisms contained in international human rights law, examining whether IoT devices can be considered technological progress and, if so, whether they are protected under the Covenant on Economic, Social and Cultural Rights. Moreover, I will present the interrelations and reciprocal limitations arising between the freedom of scientific research and the right to property, both of which are protected under the Covenant as well.

The second issue is the possibilities available researchers as the EU citizens. I will focus on two questions: first, whether the collected data falls within the scope of legislation on the re-use of public sector information, and second, whether regulations on the protection of non-personal data will apply to the data collected by IoT medical devices, enabling their re-use.

---

*Speaker

Digital collections and databases

Database of the Research Project Reframed Image: Reception of Prints in the Kingdom of Poland from the end of the Fifteenth to the Beginning of the Seventeenth Century

Grażyna Jurkowlaniec * [1], Piotr Kopszak * [2]

Michał Kozak * [3]

[1] Faculty of History, University of Warsaw – Poland
[2] National Museum in Warsaw – Poland
[3] Poznań Supercomputing and Networking Center – Poland

Research project (Grażyna Jurkowlaniec)

While the artistic production of prints in the Kingdom of Poland flourished from the beginning of the seventeenth century, various manifestations of interest in the medium can be traced back to the end of the fifteenth century. Graphic designs were adopted by artists working in different materials and techniques. Impressions were collected, but they were also instrumentally used for examples in confessional disputes and diplomatic actions. The aim of the Reframed Image project (http://reframedimage.uw.edu.pl/) is to investigate the intricate links between the intentions of the patrons, artists, and beholders (originally intended and subsequent), thus providing insight into the mechanism of dissemination and reception of prints and the agency in those processes of Poles, active both in Poland and across Europe.

The initial task of the project, which started in February 2016, was to create a working tool: an online database gathering traces of reception of European prints in the Kingdom of Poland before the beginning of the seventeenth century and all *polonica* related to the most significant European printing centers of the time. Ultimately, the database is expected to present a repertoire of printed images that had been available in the past to representatives of various artistic professions and their patrons, and thus to determine the scale and characteristics of collections of prints that once existed in Poland and reveal the place of the Polish art in the reception and formulation of the European print culture.

---

*Speaker

The database currently contains several thousand entries.These are mainly artworks created in late medieval and early modern Poland and their printed prototypes, but also include people, places and objects related to Poland and reproduced in European prints, as well as references to prints in written sources connected with Poland. The database allows one to organize, systematize, and analyze the data according to various criteria. The records are linked through carefully selected and systematically inserted key words, which facilitate categorization and various specialized or broad searches-according to printmakers, designers of the graphic models, iconography, function, chronology, and geography.

Database model (Piotr Kopszak)

The accumulation of systematically organized material, hitherto dispersed, is a point of departure for various enquiries. As a consequence, crucial problems arise: every stage of research reveals further aspects of the phenomena under inquiry. Also, the investigators constantly insist on various amendments and modifications of the model of the database in the course of the research.

Such opportunities are offered by adoption of the ISO/IEC 13250 Topic Maps standard as well as TMCL (ISO 19756, now at draft stage) that provides the rules for defining topic types. The conceptual model CIDOC has been adopted as a basic framework of specific types of topics and relationships among them, although it is not an intention of the project to implement the CIDOC model at 100%. Rather, complementing CIDOC with various types of topics or relationships proved indispensable, particularly with respect to the relationship of influence among works of visual arts. Topincs software has been chosen as a tool that enables designing both the model of the database and the user's interface.

While Topincs is used to enter the data into the constantly developed working database, the publicly accessible database (http://urus.uw.edu.pl/) is generated by Middleman: a static site generator that downloads the data through services defined in Topincs and shared in the YAML format. Since the granular model of the working database demands expertise, it has been simplified in the public database to make it user friendly.

Search engine (Michal Kozak)

Because of limited capabilities of the Topincs software in the area of searching for data, it was necessary to implement a separate solution. Our solution uses the Apache Solr search engine with faceting, which offers users advanced data filtering possibilities. Besides the basic filters such as *type of data*, *materials*, *techniques,* and *date range*, the system allows for filtering by *people* and their *roles*. For instance, it is possible to search all objects in which someone is an *engraver* and then filter by specific engravers. It is also possible to filter data first by *persons* and only then by specific *roles* of these persons (e.g. *engraver*, *purchaser*). Another non-trivial filter is *places*, which have their hierarchy in the database. Thus, filtering by *places* implies that all respective 'subplaces' are taken into account too. For example, if *Paris* is selected, all objects related to *Bibliothèque nationale de France in Paris* will also be included. Similar, but more complex, is the iconographic filter of *topics and themes*, with a more complicated hierarchy of the data structure, as each level of hierarchy can have more than one parent.

To achieve these objectives, the data from Topincs must be indexed in Apache Solr in many nested indexes. Our indexing software also uses data from YAML files downloaded from Topincs. The data indexed in Apache Solr is the only source for our search portal implemented in Java and Play Framework. This search portal is insensitive to diacritics, uses facets as described above, and offers sorting by type of object, lexicographic order, and chronological order.

GIS in gathering information from historical sources

Michał Gochna * [1], Grzegorz Myrda * [1]

[1] The Tadeusz Manteuffel Institute of History, Polish Academy of Sciences – Poland

Amongst many types of historical sources used in research of history of Poland from late medieval times to the the end of 18th century some have fortunately preserved in a large amount. Those are mostly court books and fiscal sources. Although they bring a significant information regarding many political, social and economic aspects of history, their massiveness brings major difficulties in obtaining necessary information. Many attempts have been made to simplify the access to content of those sources. In case of fiscal documents, full-text editions of tax registers and various kinds of bills and receipts has been prepared. Elaboration of court sources, which are mostly books of court records, vary from full-text editions in the case of the oldest books, to the summaries of content in the case of later books. Other way of making historical massive sources accessible, which is being more and more popular, is to prepare scans of manuscripts and put them on the Internet with an open access. An example of the latter can be found on the website "Search in Archives", where the Polish State Archives successively publish new scans of documents. In this approach, however, one can work only with a simple scan, equipped with basic metadata (storage place, signature, title of the unit). Acquiring data from these materials requires reading the source content in the same way as in the case of working with originals in the archive. In other words, two main approaches of publishing massive historical sources could be lined out. On the one hand it is a traditional, full-text edition or preparations of summarizes, both, in most cases, requiring a significant amount of time of research work; on the other hand, a simple on-line publication of the scans of manuscripts. While the first approach is valuable in publishing sources of limited size, it does not solve the problem of facilitating the use of mass sources. The publication of scans, in turn, changes only the form of access to the source and the medium on which it is presented. The problem of developing a larger portion of material in a reasonable time still last.

A proposition of solution of this issue has been developed during the preparation of the projects regarding the historical geography of Poland that were, and still are, conducted in the Tadeusz Manteuffel Institute of History of Polish Academy Of Sciences. Two types of digital editions of historical sources were prepared: a digital edition of tax registers from the Greater Poland from the 2nd half of the 16th century, and a digital edition of Greater Poland court books from 16th century. Both of them were prepared with the use of GIS (Geographical Information Systems) tools and/or GIS technology, as those can be seen as universal platform for collecting, processing, analyzing and sharing data from different sources and of different kind, including geographical, but also non-geographical information contained in historical manuscripts.

---

*Speaker

The use of standards existing in GIS world, resulting from natural desire to reduce the costs of time-consuming data gathering used by many different groups of recipients, contributes to increased interoperability when it comes to utilising collected information.

Obtained data stored in database is software agnostic, however compliant with hundreds of software products that are already implementing or have been certified with Open Geospatial Consortium standards, so can be used in wide range of applications for further processing.

According to above rules some software during different projects was developed to facilitate the process of entering specific data, like digital editing of historical manuscripts application called INDXR or OntoForm aimed at helping in building database driven by ontology.

Foundation for all of above solutions is common database, accessible through the Internet, for everyone interested in using it. The data collected can be expanded, updated, verified using dedicated web tools associated with database or any GIS application to which the user is accustomed to. This makes it much easier to achieve cross-project synergies using the same collected data, or enriching it with additional layers of information.

Work were carried out by historians, geographers, ontologists, computer scientists, who were trying to find theoretical foundations and methods of collecting source information in a way that allows integration of data descending from different sources, different periods of time, different areas, and having different content.

Historical Atlas of Poland in the 2nd half of the 16th century. Integration of cartographic data from various sources.

Tomasz Panecki [1], Katarzyna Słomska * [1], Grzegorz Myrda [1], Danuta Stepińska [1]

[1] Institute of History, Polish Academy of Sciences – Poland

The Department of Historical Atlas at The Tadeusz Manteuffel Institute of History (PAS) has been conducting a project which has been dealing with a cartographic representation of 16th century administrative landscape. Volumes of "Historical Atlas of Poland in the 2nd half of the 16th century" cover at this moment most of voivodeships of Polish Crown depicting settlements, administrative units, roads and physiography. The idea of the "Atlas" stems from the year 1880 when Stanislaw Smolka proposed elaboration of such maps on the first General Congress of Polish Historians in Cracow. Maps with settlemet network and administrative units were to be used by scholars as base maps for historical and geographical studies. The work on the "Atlas" in current format started in 1966 and is to be finished in 2020. Several years ago most of the data from previously elaborated volumes were integrated into spatial database which grows as the project develops.
Each volume includes the so-called "main map" (scale 1:250 000), thematic maps (1:500 000) and plans of the most important cities and towns (1:10 000). The scope of content remains the rather same for each volume from the beginning of the series. However, the introduction of Geographic Information Systems (GIS) to our work radically changed the whole process of historical mapping.

In our talk we would like to focus on the data integration from the "main map". It presents most of all settlement network (with names, types, mereological relations and ownership), administrative divisions (ecclesiastical and secular), most important roads and physiography (forests, rivers, swamps, relief). The goal of the paper is to present and discuss methodical problems which occur during the integration and harmonization of main maps from all previously elaborated volumes as well as these which are currently under development.

First five volumes, concerning Mazovia (voivodeships of Rawa, Plock, and Mazovian), Lesser Poland (voivodeships of Cracow, Sandomierz, and Lublin), and Sieradz and Łeczyca Voivodeships were elaborated between 1966 and 2008. In the 2014 maps of eight voivodeships from that volumes were merged and prepared for english publication. They are available online, not only in the form of static map but also in the form of the downloadable database (The map of Polish lands of the Crown in the 16th c. – a spatial database at the Atlas Fontium website www.atlasfontium.pl). The publication is also available in a paper version, in which the main map is divided into four sheets because of practical reasons.

---

*Speaker

Subsequently, maps are elaborated with the help of GIS tools, which on the one hand allowed to reduce time needed for map preparation, but on the other, introduced new problems. That concerns "Greater Poland" volume, which was published in 2017 and Kuyavia, Podlachia and Royal Prussia which are to be published in 2020. It is worth noting, that not only cartographic team of the "Atlas" works in GIS software, but digital tools have a substantial role in the work of historians involved in the project.

At the end of the project (2020) the goal is to have all previously elaborated volumes integrated into one publication both digital (spatial database) and printed. At this moment we can point out at least four different datasets which need to be merged and unified.

1. Volumes elaborated from 1966 to 2008 were integrated in 2014 using Adobe Illustrator/Photoshop and only partially converted to GIS formats.

2. "Greater Poland" volume (2017) elaborated fully in GIS environment.

3. "Kuyavia" and "Podlachia" which are currently under elaboration (also in GIS software).

4. "Royal Prussia" which were elaborated by M. Biskup (1961), but in lower scale (1:500 000) and different scope of content.

The scope of our works involved at the first stage data transformation from graphic-vector files to GIS-vector files and further georeferencing. Then, data prepared in that way had to be adjusted on the borders of voivodeships (especially between older and newer volumes) so they would not overlap or have gaps. It also involved harmonization of attributes, such as names or types of features. The most important work however still concerns harmonization of settlements and administrative units (work in progress). As the "main map" covers large amount of symbols, settlements are distinguished by their type, role, size, ownership and location precision, a database model for handling such a symbology has to be created rather than storing it in GIS application. In this way, each settlement symbol which is to be depicted on the map has its own symbol identifier in the database linked to particular entity (settlement). Those identifiers are also stored in GIS desktop application allowing to provide fast symbolization directly from the database. Furthermore, for administrative units modelling we use the so-call Least Common Geometry method (LCG) which is based on precinct data from National Register of Borders (pl: Państwowy Rejestr Granic). All historical administrative units are spatially identified with contemporary precincts. If one precinct covers more than one historical unit it has to be splitted. In this way we are actually elaborating a methodology for modelling all historical borders, not only those coming from 16th century.

As final results, we plan to prepare the main map prepared in scale 1:250,000, and a database with the settlements and physiography of the whole Polish Crown. All maps and databases will be available online and downloadable for all interested parties.

The nature of Humanities data in a case study: a digital collection of Greek epigraphic heritage from Bulgaria

Dimitar Iliev * [1]

[1] University of Sofia – Bulgaria

The present paper's point of departure is the premise that, in order to better understand the nature of data in the Humanities, the bottom-up approach is the most fruitful one. This implies case studies made for each project or initiative working on a particular set of objects (texts, images, monuments, artefacts, etc.) and applying particular DH methods and tools. Such case studies can successfully describe the specificities of processing the given objects as well as the issues and results of the given approach. When many such case studies are compared, trends and patterns can be detected and least common denominators can be drawn that enable a better understanding of what Humanities data at large are like and what is special about dealing with them in comparison with other types of datasets (e.g. scientific, statistical, etc.).

The topic of this paper will be the Telamon Project which aims at creating an online collection of the Ancient Greek inscriptions from the territory of today's Bulgaria. It was started some years ago by a team of scholars at the Department of Classics to the University of Sofia, it has received technical and financial aid from the University's IT Centre and the Centre for Excellence in the Humanities and it currently undergoes the finishing stages of its preparation in the framework of the National CLaDA-BG Consortium (conducting activities and implementing national contributions for CLARIN-EU and DARIAH-EU in Bulgaria). Its web site with a small initial collection of inscriptions will be launched for the public in 2019.

Ancient Greek epigraphic heritage in Bulgaria encompasses monuments from the period between 6. c. BCE and 4. c. CE covers the territory of the whole country and counts more than 4,500 inscriptions (the number is growing as new monuments are constantly found and published). For the time being, the project team concentrates on presenting the inscriptional heritage of only two major Greek cities in Roman Thrace, Philippopolis (today's Plovdiv) and Augusta Traiana (today's Stara Zagora). The perspectives are to gradually expand the territory covered. But even the fullest and exhaustive collection of monuments would be relatively small from the point of view of data science. The whole of ancient inscriptions from Bulgaria would still be quite far from big data. In Digital Epigraphy and the related DH fields, even large collections of texts, images, and objects from a single given project initiative are mostly amenable to processing and analysis and their data management does not necessarily require heuristic approaches.

But, in order to fully cover even the first several hundred or so monuments from Roman Thrace, a significant amount of time, effort and specialized skills need to be invested. The nature and

---

*Speaker

content of the monuments can be diverse and complex. For the description of all the features thereof, TEI XML according to the standards and schemas of the EpiDoc subset for epigraphic and papyrological documents is applied. This makes possible the thorough and detailed description of all the characteristics of the text and the material object that carries it, together with its history, its editors in digital and analogue corpora and other similar metadata. The visualisation, cataloguing and indexing of such data which is done with the EFES front-end service and editing platform depends very much on this accurate description. All this implies a great stress on the adequate encoding of single monuments and its important place in the workflow as compared to the subsequent processing which is much easier if the good quality of the encoded data is ensured.

This leads to the question of the human resources needed to conduct such work. The most valuable contributions come from trained epigraphers who have additionally acquired XML encoding skills. That is why regular trainings are essential for the sustainability of such a project. On the other hand, the cross-section of scholars skilled both in the complex discipline of Greek epigraphy and in the technicalities of mark-up will always be relatively small and few people at the same time could be recruited for such initiatives resulting in the gradual pace of the enlargement of the digital epigraphic corpora.

Crucial information conveyed by historical documents like ancient inscriptions are the names, events, people and places mentioned in them: in other words, named entities. The encoding of NE's, together with dates and periods, and their proper organising in authority lists are among the most important tasks of an EpiDoc encoder. Its importance is increased by the linking of the respective NE's with external authorities such as prosopographies and gazetteers such as Pleiades, together with the content linked to it from Pelagios, which enables the integration of monuments from a particular collection to a larger set of linked open data.

These are some of the important characteristics of a digital epigraphic collection that became evident in the work on the Telamon Project: great importance of adequate description of the (meta)data by specialists in the field; relatively small volume of the encoded and processed data in the scope of a single project; the priority of named entities, and the importance of linked open data. These features will be demonstrated with examples from the team's practice and the inscriptional material we have at our disposal in search for connections and comparisons with other initiatives in order to draw more general conclusions for the nature of data in the Humanities.

Keywords: digital epigrpaphy, TEI, XML, EpiDoc, CLaDA, BG, Greek inscriptions, Bulgaria

"How similar, how different? The problems of quantification of similarity in archaeological research- the example of ancient pottery "

Agnieszka Kaliszewska * [1], Rafał Bieńkowski [1], Monika Syga [2]

[1] Systems Research Institute [Warsaw] – Poland
[2] Faculty of Mathematics and Information Science [Warszawa] – Poland

This contribution aims at presenting issues arising with recognition and description of similarity in the archaeological research. This topic might seem trivial at the beginning but it can be the source of numerous difficulties in the research process. This is in major part related to the fact that each person, even a theoretically objective researcher, perceives reality in a slightly different way. These issues have been partly addressed in archaeology and geological sciences with tools such as the Munsell colour system or the Moh scale. The necessity to use objective and unequivocal parameters and terminology in the description of the subjects of the study is one of the big challenges in archaeological research.

For the purpose of this study, we use the example of archaeological pottery. Pottery is the most common and numerous find during archaeological excavations, and hence constitutes a good marker for chronological periods, cultural groups, sometimes even ethnicity of the people who created or used the pots. The study of ancient pottery requires a careful study of the similarities and differences between the pots in one assemblage as well their relation to the pottery found elsewhere. Oftentimes the definition of whether two objects are different or similar and to what degree poses a challenge. The differences can often be very subtle and the recognition of the right parameters that distinguish the pottery fragments requires highly specialised knowledge.

The research on archaeology is heavily based on typologies, created mostly on the base of the shape and sometimes decoration of the vessels. Typologies constitute an efficient tool to deal with the variability of ancient ceramics. They can easily be referenced and we can assign a "new" vessel to one of the known groups. However, researchers often encounter the situation when they want to quantify how much two vessels are similar. This, in turn, requires a definition of similarity as well as a way to measure it. The easiest definition of similarity is that it is a set of characteristics shared by the objects. Among these characteristics are size, weight, capacity, or the material of which the object is made. A commonly used characteristic is the shape of the object. Although it is very natural to describe an object by its shape, it creates a series of problems when trying to compare two objects.

In archaeological research, there are several standards of conveying as many information as possible about pottery and pottery fragments. These are used for the purpose of preparing data

---

*Speaker

for analysis and to enable comparison between fragments. Among these are: archaeological drawings or a standardised form for the description of specific aspects of pottery. With the introduction of digital photography and microscope photography, this kind of documentation is also often added. The need to create highly precise and complete documentation is a part of the whole archaeological research, not only ceramology.

The process, described above is true for an "analogue" form of research. Many problems arise when the process is transferred to a digital environment. In order to perform any kind of analysis digital data must be complete, consistent and compatible. If we are to look for similarity/differences in the investigated objects, similarity must be well defined, and, contrary to the analogue process there is no place for judgement calls made on the basis of experience or knowledge. The second issue which arises when transferring the analysis to the digital environment is the question of compatibility of data. If the documentation is in the form of descriptions, there is no guarantee that characteristics perceived as similar are described in a similar way. Such inconsistencies may include terms such as ovoid, oval, cylindrical, rounded-elongated. To further make things more complicated these terms are also used together with quantifiers such as slightly, a little, sharply, very etc. All these expressions, although very natural in written language, do not translate in the computer-aided analysis. Many of these can be tackled through tagging of data, dictionaries or ontologies, but the question of "How similar are the objects?" will be still dependent on the adopted definition and measure of similarity. Moreover, such definition and measure will differ from study to study, as they depend on the research questions that need to be answered through the analysis.

In the last part, the authors would like to present an example of an approach to determine similarity and automatically create typologies in archaeological ceramics through the application of mathematical tools and examine how it relates to the discussion presented in the contribution.

Keywords: similarity, typologies, archaeology, data analysis

Infrastructures and beyond

DH Foresight: informing the future of methods and infrastructures

Mark Hedges * [1], David Stuart * [2]

George Tzedopoulos [3], Vicky Garnett [4], Sheena Bassett * [5]

Claus Spiecker * [6]

Maurizio Sanesi [7], Roberta Giacomi [7]

[1] Kingś College London – United Kingdom
[2] King's College London – United Kingdom
[3] Academy of Athens – Greece
[4] Trinity College Dublin – Ireland
[5] PIN – Italy
[6] FH Potsdam – Germany
[7] SISMEL – Italy

Overview

In recent years there has been rapid growth both in the development of digital methods and tools and in their application across a wide range of disciplines within the humanities and cultural heritage studies. In parallel to these developments, there have been numerous initiatives and projects at both national and international levels dedicated to the creation of more coordinated research infrastructures at different levels of subject specificity. These projects have been marshalling and integrating data, tools, services, technologies, policies and human resources in support of advanced research in universities, cultural heritage institutions, and other organisations. The future development of this landscape depends on a complex and dynamic ecosystem of interactions between: changing scholarly priorities, questions and methods; technological advances and new tool development; and the broader social, cultural and economic contexts within which both scholars and infrastructures are situated. A sound knowledge base is required if policy-making bodies are to 'optimise' outcomes through implementing appropriate research and innovation policies, setting research priorities, and influencing the progress of research through funding programmes and other interventions.

To this end, the Horizon 2020 project PARTHENOS has carried out a *foresight study*, investigating how (digital) research methods, technologies and infrastructures in digital humanities and cultural heritage may develop in the short to medium term. It is important to understand that foresight research is not simply 'future gazing', nor is it about forecasting by experts (although experts may, and indeed should, participate). Rather, it is a way of facilitating structured thinking and debate about long-term issues and developments, and of broadening participation

---

*Speaker

in this process of thinking and debate, to create a shared understanding about possible futures and to enable them to be shaped or influenced.

A key component is the participative aspect. The vision is not that of a small number of experts, but is based on engagement with and involvement of a broad range of key stakeholders, including decision- and policy-makers, but also members of the broader community, including scholars, potential users of research infrastructures, and practitioner stakeholders such as infrastructure providers, data curators, and archivists.

The aim of this study was thus not *simply* to identify trends and to predict future evolution within the sector, but rather to enable the community to inform and influence this evolution by identifying research and funding strategies and interventions that can be taken forward by the various stakeholders active in the (digital) humanities landscape, including universities, research institutions, funding agencies, and research infrastructure providers. This study will thus feed into strategic R&D thinking within the European Commision, other funding bodies, and research organisations, and will give participants the opportunity to make their opinions known and to influence these strategic developments over the coming years, and thus to maximise the innovative potential of digital research in the humanities.

Format of the Study

Over the last year, the foresight study team has organised a series of structured, interactive workshops to obtain input for study study by curating a multi-polar discussion between representatives from various EU research infrastructure initiatives, and a range of actual or potential stakeholders in those infrastructures, including (but not restricted to) user/researchers. Engaging a representative range of relevant and informed stakeholders in the dialogue extended the breadth and depth of the knowledge base created by the foresight process, by drawing on distributed knowledge (different stakeholders having access to different information), and thus enriches and improves the decisions that may ultimately be made on the basis of our work. These events were supplemented by a series of interviews with key actors representing different career backgrounds and levels.

We framed the discussion of 'foresight' in this study by asking the following questions, and identifying:

- current trends – what is happening, and what impact is it having?
- potentialities and opportunities – what may happen?
- requirements – what do we want to happen?
- obstacles, constraints, risks and threats – what might prevent this from happening?
- what activities and interventions (e.g. funding programmes, strategic research, service provision) might serve to 'optimise' outcomes?

These trends etc. may have aspects relating to technology, to scholarly or professional practice, or to the broader 'environment' (social, cultural, economic, political, etc.), or some combination of the three. The data gathered was encoded to produce a 'knowledge base' of data about DH foresight.

The Proposed Meeting

The meeting will be introduced by a number of brief presentations, but the aim of the meeting is to be interactive, with discussion opened up to the audience. The chairs will ensure that the audience is encouraged to contribute actively, and that an inclusive environment is created in which multiple viewpoints and opinions may be canvassed. Depending on the number of attendees, the chairs may decide to organise a number of parallel break-out groups. The meeting will address the foresight study from several angles:

- **Dissemination**: Presenting the results of the foresight study to date [Note: this is of particular relevance to the event, as the proposed meeting follows on from the foresight workshop organised at the DARIAH event in Paris in 2018].

- **The Hub**: Demonstrating the use of the PARTHENOS Hub, using the Foresight Issue as a case study, both as a dissemination channel and as part of the sustainability agenda for the study [Note: the Hub is a concept that was originally proposed by the DARIAH project *Humanities at Scale*].

- **Feeding back**: capturing feedback on the results of the study, and contributing by open discussion and break-out groups to their further development and refining.

- **Sustainability**: Setting up the foresight study as a dynamic and sustainable body of knowledge. Foresight does not end at the formal end of the study, but is conceptualised as an ongoing process, not only in terms of dissemination by continual monitoring the changing landscape and updating the results of the study. The Hub will be a locus for capturing this information, but in this (and other) meetings over the next six months we intend to initiate the creation of sustainable networks of foresight research by embedding stakeholders with 'foresight interests' within research and education organisations.

Keywords: foresight, data, infrastructures

The DARIAH-EU and OpenAIRE collaboration on Open Science publishing practices

Alessia Bardi [1], Miriam Baglioni [*] [1], Pedro Principe [2], Erzsébet Tóth-Czifra [3], André Vieira [2]

[1] Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Italy
[2] University of Minho [Braga] – Portugal
[3] DARIAH – Germany

In the digital research ecosystem, we no longer produce only papers. Research results can take many forms that coexist with the research paper, such as data, software, images, teaching materials, videos and more. The possibility of sharing, disseminating, and re-using digital research products of different type is fundamental for an effective implementation of the Open Science (OS) paradigm [1,2]. OS is a set of practices of science, advocated by all scientific/scholarly communication stakeholders (i.e., research funders, research and academic organizations, and researchers), according to which the research activities and all the products they generate should be freely available, under terms that enable their findability, accessibility, re-use, and re-distribution [3].

In the current scholarly communication ecosystem, it is not an easy task to apply Open Science principles and to keep the diverse types of research outputs visible, easily discoverable and connected. In order to support the shift towards Open Science, the OpenAIRE initiative introduces and implements the concept of Open Science as-a-Service (OSaaS) on top of the existing OpenAIRE infrastructure (www.openaire.eu), by delivering out-of-the-box, on-demand deployable tools in support of Open Science.

OpenAIRE materializes an open, de-duplicated, participatory metadata research graph of inter-linked scientific products (including research papers, datasets and software), with Open Access information, linked to funding information, research communities and infrastructures by collecting more than 100 millions of metadata records from more than 9,000 scholarly data sources world-wide. Thanks to the availability of the research graph, OpenAIRE is in the position to offer added-value services to research communities and infrastructures.

In the context of the EC H2020 OpenAIRE-Advance project, DARIAH-EU and OpenAIRE are collaborating on this topic and piloting one of those added-value services: the Research Community Dashboard (RCD).

The OpenAIRE RCD (https://connect.openaire.eu) provides Research Infrastructures (RIs) and its researchers with the services required to connect the research life-cycle the RIs support with

---

[*]Speaker

the scholarly communication infrastructure, in such a way science is reusable, reproducible, and transparently assessable.

The OpenAIRE RCD allows the easy deployment of community-specific gateways that work as single access points to a virtual space including metadata descriptions of all products in the OpenAIRE research graph that are relevant to the community (e.g. articles, datasets, software, workflows) as well as links between such products. Via the community gateway, researchers can also find Open Science-oriented publishing tools that they can use to manually deposit research products of any kind via Zenodo (www.zenodo.org), the OpenAIRE "catch-all" repository run by CERN, specifically thought for researchers without an institutional/thematic repository of reference, but that can be used by any researcher willing to publish any type of research product. To group depositions of the same community of practice, researchers can use the mechanism of the "Zenodo communities", which can be considered "containers" of research products related to a specific topic. When a gateway is deployed for a given community, a dedicated Zenodo community is also created.

A community gateway is managed by a set of RCD managers, which can configure the criteria that the OpenAIRE algorithms will apply to the graph for the identification of relevant products. Specifically, RCD managers can configure the algorithms by choosing relevant projects, content providers (e.g. institutional repositories, data archives), and Zenodo communities and by specifying the acknowledgment statements to be found in the full-texts of Open Access publications. Every product linked to the selected objects in the OpenAIRE graph will be automatically marked as relevant for the community (i.e. every research result "funded by" a selected project will appear in the community gateway).

Researchers can also contribute to enrich the graph by *(i)* depositing their products on Zenodo, *(ii)* assert (i.e. claim) that an existing product is relevant for the given community, and *(iii)* assert links between products. As a consequence, researchers populate and access an open, participatory scholarly communication graph of interlinked objects dedicated to their RI, through which they can share any kind of products in their community, maximise re-use and reproducibility of science, and outreach the scholarly communication at large.

The graph is also analyzed for the calculation of statistics that support the monitoring of research impact of the RI and the uptake of Open Science principles among its researchers.

The RCD for DARIAH-EU is currently available as a beta at https://beta.dariah.openaire.eu/ and the DARIAH team is currently configuring it in order to instruct OpenAIRE algorithms about the criteria by which research products available in OpenAIRE should be assigned to the DARIAH-EU RCD. In particular, OpenAIRE algorithms are being configured to find research outcomes produced thanks to the DARIAH infrastructure based on:

Acknowledgment statements available in the full-texts of Open Access publications;

Links from research products to DARIAH specific projects (e.g. DESIR, HaS-DARIAH);

The provider from which OpenAIRE collected the metadata records (e.g. all products from the DARIAH collection of HAL);

Relevant Zenodo communities (e.g. the DiMPO Zenodo community).

Once the configuration is stable and applied by OpenAIRE algorithms, the RCD will show relevant statistics about the uptake of Open Science practices, such as the percentage of Open

Access publications and numbers of publications linked to datasets and to software, and the research impact of DARIAH. The DARIAH team will collaborate jointly with the OpenAIRE team for the definition of the indicators to be calculated, possibly based on the request and expectation of the DARIAH management and funding agencies.

References

1. European Commission. Validation of the results of the public consultation on Science 2.0: Science in Transition [report]. Brussels: European Commission, Directorate-General for Research and Innovation. (2015). http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf.

2. European Commission's Directorate-General for Research & Innovation (RTD). Open Innovation, Open Science and Open to the World. (2016). https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe.

3. FOSTER. Open Science Definition: https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition.

Sustainability beyond guidelines

Andrea Scharnhorst * [1,2], Francesca Morselli * [1,2]

Jennifer Edmond * [2,3]

[1] Data Archiving and Networked Services – Netherlands
[2] Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France
[3] Trinity College Dublin – Ireland

SUSTAINABILITY BEYOND GUIDELINES
How our perception of research infrastructures shapes our understanding of sustainability?
This paper addresses the question of the human factor or human capital when reflecting about
the sustainability of research infrastructures. In the heroic endeavour to build pan-European
research infrastructures one could gain the impression that attention goes primarily to the pro-
cess of building them, and less to those which execute this process. If we observe the relevant
discourses about infrastructures, that is the science policy discourse and science and technology
studies reflection, we see that humans - people - indeed do figure in them. They appear in
their roles as 'staff' (working in RIS) and 'users' (using the RIS), or as careers or knowledge
of different types inherent to making the infrastructure. Our paper aims to contribute to the
awareness raising process about the roles of actors in and around a research infrastructure and
their function for sustainability.

We start from the results of two projects: CENDARI, a research infrastructure project which
looked into its own sustainability; and Humanities at Scale (HaS) which devised a so-called
reference or architecture model to organise and communicate about the achievements (contri-
butions) made in an infrastructure and the different roles and functions of those involved in
it. From the thorough analysis of CENDARI about sustainability we learn about the process
character of sustainability and the key position people (as stakeholders, as co-creators, as users)
have to ensure sustainability in a continuous and ongoing way. In a complementary way, the
HaS reference model unfolds the complexity of processes in a distributed work environment,
the different roles, skills and competences which are required to execute them, and delivers a
systematics to describe them.

With evidence for the importance of human capital and knowledge capital for sustainability
and toolsets to describe the required work processes, we arrive at the question how to sustain
the people in order to sustain infrastructures.

We conclude the paper with reflections how the careers of those working in and with infras-
tructures can be fostered by means of formal education and training, and how their functions

---

*Speaker

can be acknowledged by occupational classifications.

References

CENDARI Project, http://www.cendari.eu

Christine L Borgman, Paul N. Edwards, Steven J. Jackson, Melissa K. Chalmers, et al.. "Knowledge Infrastructures: Intellectual Frameworks and Research Challenges" Deep Blue(2013) Available at: http://works.bepress.com/borgman/318/

DARIAH Reference Architecture, https://priddy.github.io/DARIAH-RA/service/

Humanities at Scale Project, http://has.dariah.eu/

RM-ODP: The ISO Reference Model for Open Distributed Processing, Antonio Vallecillohttp://ae.aim.nl/weblo
Sustainable European Research Infrastructures - A call for action. COMMISSION STAFF WORKING DOCUMENT. Long-term sustainability of Research Infrastructures - SWD(2017) 323 final

Visual Analytics of the DARIAH in-kind contributions

Andrea Scharnhorst * [1,2], Femmy Admiraal * [1,2]

Dirk Roorda [1]

[1] Data Archiving and Networked Services – Netherlands
[2] Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

One of the main products of the Humanities at Scale project was a profound (re)-conceptualisation of the DARIAH in-kinds and a web-based service [2] to collect, review and display in-kind contributions, in short DARIAH contributions. [1] This paper looks at the results of the implementation phase of this new service, and particularly focuses on baseline statistics and visual analytics of the content submitted (about 300 contributions for 2017 and 2018).

In general, for European Research Infrastructures, so-called in-kind contributions are a way for the members to account for their national efforts under the umbrella of the ERIC. They may represent contributions available for all ERIC members (e.g., central services executed by an institution in a member country) and/or contributions which embody, complement, or enhance the mission and strategic actions of an ERIC on the national level.

DARIAH's reference model [1] on the basis of which contributions are defined, introduces two main categories: 'services' and 'activities'. For them a detailed metadata scheme has been devised. Submitted contributions are further subject to a detailed self-assessment and reviewing process, one part of which is dedicated to determine those contributions which are put up for the financial accountability of a member's contributions.

The web-based service replaces earlier forms of template-based and data-based submission of in-kinds, and enables immediate comparison of the submissions - also by a couple of visual interfaces (map, tables).

In this paper, we aim to demonstrate the main benefits of the tool and zoom into three aspects.

The DARIAH contribution tool relies on the tedious and comprehensive work of the National Coordinators which are in charge of the submission process. To make their often invisible work more visible is one motivation behind this paper.

---

*Speaker

The submitted content as such forms an interesting empirical base for reflection on what is seen as a DARIAH contribution by the DARIAH community. This, in turn, can inform the DARIAH strategy and help to monitor the success of its actions. This is the second motivation.

Submissions can still vary in form greatly, despite the formal model. They remain human generated content. Analysing the variety helps to tune the tool to user requirements, to fix bugs, and to curate the data it collects.

We conclude our paper with reflections on the future use of the tool and its connection to other DARIAH strategic actions, as designed in the Strategic Action Plan II.

PS: We would like to also present a poster on this topic.

D5.1 Report on Integrated Service!Needs: DARIAH (in kind) contributions - Concept and Procedures Lisa De Leeuw Femmy Admiraal Matej Ďurčo Nicolas Larousse Michael Mertens Francesca Morselli Mike Priddy Paulin Ribbe Carsten Thiel Lars Wieneke https://hal.archives-ouvertes.fr/hal-01628733
https://dariah-beta.dans.knaw.nl/data/contrib/filter

Keywords: visualisation, contributions, DARIAH services, DARIAH STRAPL

Archiving and enriching data

Linking MIDI Data for multidisciplinary and FAIR AI on Music

Albert Meroño Peñuela * 1

1 Vrije Universiteit Amsterdam [Amsterdam] – Netherlands

More and more researchers from a large variety of fields are using modern AI techniques to understand, describe, and generate musical content and metadata. In the last years we have seen a blooming diversity of algorithms, mainly learned automatically from data using deep learning architectures, to address various tasks in Music Information Retrieval such as instrument separation, automatic transcription, and chord detection. Some of these architectures use the models they learn to generate genuinely new, but still data-inspired, songs and tunes that start posing questions about machine vs human musical creativity. In more musicological and humanistic settings, fundamental questions and hypotheses are typically multidisciplinary in nature, normalling dealing with aspects of culture, economy, and musical traits and characteristics. The common requirement for methods in all these fields is the same: large, carefully curated, and explicitly and semantically described datasets.

Generating these datasets in such a way that they can be reused among different asks in those various fields is challenging. There exist several reasons for this: (1) existing musical data is distributed and scattered all over the Web; (2) meaningful connections between these data are missing (e.g. a Spotify song URI and its corresponding URI entry in MusicBrainz); (3) equivalent representations of musical knowledge, like symbolic music scores, are encoded in various formats (MIDI, MusicXML, MEI, LillyPond, ABC, etc.) that are hardly interoperable.

Solving these issues at scale for arbitrary data (i.e. not just musical information) has been one the driving goals of 20 years of Semantic Web research [6]. This research has reached now maturity and delivered technologies to make Knowledge Graphs possible; these Knowledge Graphs, when deployed, integrate, interlink, and make datasets interoperable on the Web. In this poster, we will reflect on current work towards the construction of Music Knowledge Graphs that integrate symbolic musical content in the MIDI format and metadata databases that describe and enrich that content. In the first of such works, we proposed the midi2rdf algorithm [1]; a program that can lossless convert any individual MIDI file into the interoperable Resource Description Framework (RDF) format through a low-level ontology. Following up on this, we used community-based knowledge from GitHub to gather a comprehensive list of sites containing MIDI files on the Web (around 500K); and run the midi2rdf algorithm on them to create the MIDI Linked Data Cloud, a MIDI Knowledge Graph of 10B semantic statements [2]. The first creative use of such integrated musical knowledge space was to enable large scale mashups (mixes of different tracks of various songs), in which the compatibility of different candidate tracks for the mix (e.g. their rhythmic similarity) is automatically determined by a SPARQL pattern [3] by the so-called "SPARQL-DJ". Building on recent deep learning architectures, the

---

* Speaker

Text2MIDI Variational Auto Encoder [5] learns the relationship between words in lyrics and melodic motifs, generating melodies automatically for arbitrary input poems.

Three challenging tasks remain open for the future. The first is to envisage algorithms for automatically linking MIDI musical content –or any other musical symbolic format– to musical metadata databases (e.g. MusicBrainz) at scale, in order to enable a richer discoverability and retrieval of related resources. First steps towards this goal are already made in the Semantic Web MIDI Tape [4] through amateur playback, MIDI similarity algorithms, and metadata propagation. The second challenge is to expand the interoperability of Linked Data Musical dataspaces to other symbolic notation formats, like ABC, LillyPond, MusicXML or MEI, which remain hardly interoperable. The third challenge is to better understand the integration requirements from the various fields doing research in AI and Music in order to conceive appropriate strategies, data models, and access methods for Music Knowledge Graphs.

References

Albert Meroño-Peñuela, Rinke Hoekstra. "The Song Remains the Same: Lossless Conversion and Streaming of MIDI to RDF and Back". In: 13th Extended Semantic Web Conference (ESWC 2016), posters and demos track. May 29th - June 2nd, Heraklion, Crete, Greece (2016).

Albert Meroño-Peñuela, Rinke Hoekstra, Aldo Gangemi, Peter Bloem, Reinier de Valk, Bas Stringer, Berit Janssen, Victor de Boer, Alo Allik, Stefan Schlobach, Kevin Page. "The MIDI Linked Data Cloud". In: The Semantic Web – ISWC 2017, 16th International Semantic Web Conference. Lecture Notes in Computer Science, vol 10587, pp. 156-164 (2017).

Rick Meerwaldt, Albert Meroño-Peñuela, Stefan Schlobach. "Mixing Music as Linked Data: SPARQL-based MIDI Mashups". In: Proceedings of the 2nd Workshop on Humanities in the SEmantic web (WHiSe 2017). ISWC 2017, October 22nd, Vienna, Austria (2017).

Albert Meroño-Peñuela, Reinier de Valk, Enrico Daga, Marilena Daquino, Anna Kent-Muller. "The Semantic Web MIDI Tape: An Interface for Interlinking MIDI and Context Metadata". In: Workshop on Semantic Applications for Audio and Music, ISWC 2018. 9th October 2018, Monterey, California, USA (2018).

Roderick van der Weerdt. "Generating Music from Text: Mapping Embeddings to a VAE's Latent Space". MSc thesis, Universiteit van Amsterdam (2018).

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." Scientific american 284.5 (2001): 28-37.

The Battleground of Humanities Data: Handling Multidisciplinarity and Diversity

Athanasios Karasimos * [1], Yorgos Tzedopoulos * [1]

Helen Goulis [1], Eirini Mergoupi-Savaidou [1], Gerasimos Chrysovitsanos [1]
Athina Iakovidou [1], Anastasia Falierou [1], Patritsia Kalafata [1], Eleni
Vernardaki [1], Maria Spiliotopoulou [1], Helen Katsiadakis [1]

[1] Academy of Athens, DARIAH-GR/DYAS – Greece

Humanities and Arts data vary significantly as to their degree of formalization. Whereas language studies and archaeology have developed formal vocabularies and/or more or less strict typologies, historical data present a particular challenge as to their spatio-temporal accuracy (fuzzy data) and as to their interpretation. Historical concepts and terminologies tend to porousness and relativism in relation to social and cultural contexts over time and space. Besides data records of the past in physical form usually contain errors and gaps, if they are not entirely missing or destroyed.The continuous effort to digitize collections of cultural heritage and Humanities data has increased the availability of and access to diverse, multidisciplinary, significant and rare materials; however, this overproduction of digital and digitized data can be easily lost when misidentified or when not meta-identified at all. Unfortunately, 80% of the data are accompanied by zero or false metadata. As digital collections grow, it is imperative that every digitized object is tagged with accurate metadata information; more specifically, all Digital Humanities data should be described by a dominant and widely accepted metadata schema to be easily searched, categorized and processed. At the end of the day, metadata are simply the vehicle to help researchers find, access and reuse data.The Arts and Humanities landscape, a record of human condition and its social, cultural and intellectual evolution, covers diverse disciplines, themselves with a long history over the ages. The common ground among disciplines is an intersecting area of new knowledge, since one of the emerging trends of DH is to address, analyze and process data through the prism of multidisciplinarity. According to Richardson (2013), the visualization of citation links in the Arts and Humanities publications shows a cluster of interrelated topics: different facets of investigation that build a complex but straightforward network of multidisciplinarity, which cannot be ignored by any research approach and analysis.

The team of the Greek national infrastructure for the Arts and Humanities (DYAS) faced the multidisciplinarity challenge when compiling a registry of all Greek collections, whether physical or digital, that are of interest to humanities and social sciences. The registry includes collections of interest to more than 15 disciplines such as Anthropology and Ethnology, Archaeology, Classics, Numismatics, Epigraphy, History, History and Philosophy of Science, Literature, Linguistics, Philosophy, Theatre Studies, Musicology, Byzantine Studies, Ottoman and Turkish Studies, Arts and Art History. The majority of information on collections is recorded in co-

---

*Speaker

operation with the curating institutions. This has often led them to reassess their method of documentation in the light of interaction with the DH infrastructure. The DYAS Collections Registry enables researchers to detect information about their research in numerous types of collections, varying from collections of objects regarding different branches of Humanities (museum collections, costume collections, carte-postales, maps, etc.) to collections of digitized texts (collections of books, manuscripts, textbooks, correspondence, written and spoken data etc). The vast amount of data and the variety of disciplines required a highly elaborated metadata schema accommodating the description of both physical and digital collections.

The DYAS multidisciplinary team has also developed the DYAS Humanities Thesaurus comprising most of the Arts and Humanities subject fields. As every discipline uses a different terminology, one of the main challenges of this effort was to agree on a common, top level coherent and consistent "backbone" Thesaurus containing a limited number of concepts from which all the thematic vocabularies and terminologies would inherit their attributes. The compilation of this discipline-agnostic Backbone Thesaurus was a bottom-up process; top-level concepts were developed by adequate abstraction from existing terminologies, meeting the demands for inter-subjective and interdisciplinary validity. Besides, the DYAS Collections Registry has also been a source for some of the terms for the Humanities Thesaurus. The terms used for the thematic description of collections, in other words a flat vocabulary, are being modelled in a structured way under the top-level concepts of the Backbone Thesaurus.

Handling humanities data is not a simple task, even within the framework of one discipline. Important issues arise concerning the diversity of the relevant sources, the institutions that curate and preserve them, and the users' requirements. Data integration in major pools creates a demand for interoperability. For example, we are currently developing a digital platform for the integration of metadata related to digitized documents, photographs, artefacts, and moving images from the 1940s in Greece. The project aims at devising an umbrella of "meta-metadata" that can bridge the varied specific metadata created by the holding institutions. This interoperable platform will enable the public and the researchers to find, access, and reuse effectively the relevant data, as well as to discover underlying connections between them.

The three aforementioned services are inscribed within an Open Science strategy, which aims at enabling new lines of inquiry by combining available sources on data. Following the necessary steps of processing for easy use (documented, machine-readable, understandable and readable in convenient format), the connection between open data and proper standardized metadata is required since this is the European Community goal to optimize the future impact of research in Europe (and worldwide). To this end, we try to reconcile the necessary standardization requirements with the complexity and diversity of Humanities data.

Bibliography

Doerr, M., Daskalaki, M., Bekiari, Ch. (2014). Thesaurus Building. DYAS, FORTH. Retrieved from

https://www.backbonethesaurus.eu/sites/default/files/ppt_thesuarus%20building.pdf

Doerr, M., Katsiadakis, H., Daskalaki, M., Goulis, H. (2015). Backbone thesaurus: Methodological outlines. DYAS, Academy of Athens & FORTH. Retrieved from

https://www.backbonethesaurus.eu/sites/default/files/ppt_methodoligical%20outlines.pdf

Karasimos, ., Goulis, H., Kalafata, P., Lipke, A., Papadakis, N., Mergoupi-Savaidou, E., Tze-

dopoulos, Y., Falierou, A., Chrysovitsanos, G., Vernardaki, E., Spiliotopoulou, M. & H. Katsi-adakis (in press). From DARIAH-GR/DYAS to PARTHENOS-EU Project: the digital infrastructures, the Collections Registries and the Thesauri in Humanities (in Greek). In Proceedings for Digital Humanities in Greece: issues and challenges. Athens: Research for the Humanities (RCH).

Richardson, M. (2013). Mapping the Multidisciplinarity of the Arts & Humanities. Research Trends 32.Elsevier Publications. Retrieved from https://www.researchtrends.com/issue-32-march-2013/mapping-the-multidisciplinarity-of-the-arts-humanities-2/.

Wineburg, Samuel (1991). On the Reading of Historical Texts: Note on the Breach Between School and Academy. American Educational Research Journal 28 (1991), pp. 495-519. Retrieved from

https://pdfs.semanticscholar.org/6499/13b351565c07d5d53f852b954eac09a1b0da.pdf?_ga=2.124481870.511527 174633602.1550749431

Finding listening experiences in books

Enrico Daga [*] [1], Enrico Motta [1]

[1] The Open University – United Kingdom

The Listening Experience Database Project (LED) is an initiative aimed at collecting accounts of people's private experiences of listening to music. Since 2012, the LED community explored a wide variety of sources, collecting over 10.000 unique experiences. In this work we take the case of supporting users on finding evidence of listening experiences in books. Particularly, we report on the approach that guided the development of a system that traverses the content of digitised texts in search for passages remarking a description of a musical event - an account of an experience of listening to music. We experimented with several approaches involving Statistical NLP and Machine Learning techniques and compared them against a curated gold standard. The best performing method has been used to develop a novel tool to support curators in discovering new listening experiences. FindLEr analyses the content of books for relevant paragraphs systematically, thus reducing significantly the cost of finding candidate listening experiences.
(extended, 1000 words, abstract attached as PDF)

Keywords: Information Retrieval, Listening Experience, Statistical NLP, Machine Learning, Music

---

[*]Speaker

Fixity in digital archive

Mikolaj Machowski [*] [1]

[1] National Museum in Warsaw – Poland

Fixity is the state of being unchanged or permanent.
Paper will discuss problem of fixity in relation to archives of digital photos.

National Digital Stewardship Alliance, consortium of over 200 institutions
specialized in preservation of digital information suggests 4 levels of data
integrity techniques.

1 level: check file fixity on ingest and create if it wasn't provided
2 level: check fixity on all ingests, use write-blockers when working with original media, virus
checking of high-risk content
3 level: check fixity on content in regular intervals, maintain logs, ability to detect corrupt data,
virus checking of all content
4 level: check fixity on specific events, ability to replace/repair corrupted data, no one should
have write access to all copies

Paper will discuss how to implement them with regard to fixity and archive
integrity in examples how it works in real life archive and how it could be
implemented in ideal world.

Base method to test fixity are checksums. Various algorithms will be discussed
with their relation to security (which ones should be avoided), economics,
convenience.

- MD5 and why you must not ever use it in anything related to security
- SHA1 and why it should be avoided
- SHA2 and its various levels

Paper will suggest way how to store (and test) checksums inside of file without
changing of checksum of visual information. This method and its variants could
be also used for fixity of audio files.
Paper will present how various changes in file reflect on checksum value.

Keywords: fixity, digital archive, visual information, checksums

---

[*]Speaker

Challenges of Digital Data for Linguistics: New Units and Hidden Processes

Manuel Alcantara-Pla * [1]

[1] Universidad Autonoma de Madrid – Spain

Internet has provided us with an amount of data without precedents. For those who research in Linguistics, it is an unexpected gift with a huge potential. The most obvious difference of the Internet discourses -also known as Computer Mediated Communication (CMC) discourses- with the previous ones is that every interaction is registered and can be retrieved. Unlike other data, texts have not suffered a typical process of digitalization (loosing information). Digital texts can be easily stored, copied, and modified. It makes communication more complex because of phenomena such as remediation, but it also gives us a great opportunity for studying discourse on a large scale.

However this gift comes with important challenges that we have to face. I focus here on two specially relevant for data analysis: the definition of units of analysis and the influence of the platforms we get the data from. Though the cases will be taken from the discourse analysis discipline, both challenges are general to Digital Humanities.

Social media and mobile phones are two inventions that have radically changed our relation with the Internet. Users are (and are requested to be) "more active, participatory, and collaborative" (Heyd 2016, 90). Many new communication genres have been created in the last decades (blogs, wikis, social networks, and micro-blogs). All of them are multimedia, and texts compete in importance with images, videos, and sounds. The most recent platforms (such as YouTube and Instagram) explicitly relegate texts to a secondary role.

When we call "chats" to interactions made up of written texts, videos, and images, we are using this term in a very broad sense. Most of the mentioned genres are asynchronous, even those more conversational. Interaction strategies are not the same as in spoken conversations, they are written or multimodal instead of spoken, and they follow new politeness rules. Therefore they do not correspond with traditional conversations, and the definition of the units of analysis of these new genres is still work in progress (Alcántara-Plá 2014).

The work of defining these new units has to meet two basic requirements. On the one hand, being multimodal communication, texts can not be analyzed independently. Words are interconnected with images and videos, and they must be studied as such. Limits between different modes have become fuzzy, and the different disciplines that are focused on them should take it into account.

On the other hand, boundaries have also become fuzzy in another sense. Spoken conversations can be divided into smaller units, from the conversation itself to turns, utterances, and

---

*Speaker

lexical units. However, these units are difficult to delimit in the Internet. Digital conversations can start in a platform and continue in a different one: a tweet might end up as part of a Whatsapp interaction or embedded in a video in YouTube. Digital information is very easily remediated, and it is a frequent habit in the Internet.

Regarding the second challenge we face when using digital data for research in Humanities, the design of the platforms we get the data from should be taken into account in our analysis. This design is a key element in the context of any mediated communication, and rarely acknowledged as such. The main problem here is that most designs are not open and they can only be interpreted analyzing their user's behavior. It is not possible to analyze them as independent variables. In order to have the most reliable information, collaboration with the owners of the platforms and their designers is crucial.

Taking the discourse analysis as example, the mentioned characteristics (its units, politeness, multimodality...) are not free choices made by the users, but affordances of the platforms where the communication is taking place. The technological mediation of the digital communication and its hidden processes determines the language we use when communicating in digital contexts. Therefore affordances and restrictions of every platform should be part of the data we use for our research.

Keywords: CMC, digital data, units, analysis, Linguistics

Technical solutions

Pathfinder: a system for data exploration, analysis, and visualization of art catalogs and exhibitions

Nuria Rodríguez-Ortega * [1], Antonio Cruces Rodríguez * [1]

[1] UNIVERSIDAD DE MÁLAGA – Spain

The growing consolidation of art exhibition studies as a research field has been accompanied by the emergence of numerous projects devoted to the construction of databases and archives, as well as the development of digital platforms for consultation and analysis. The Exhibitium Project (www.exhibitium.com) falls within this research framework. Coordinated by the iArtHis_Lab research group of the University of Málaga (Spain) since 2015, in collaboration with various international partners, one of the objectives pursued by this project is to design and develop technological infrastructures with the aim of enhancing research on art exhibitions as well as opening new critical inquiries (Rodríguez-Ortega, N. and Cruces Rodríguez, A., "Development of Technological Ecosystems for Cultural Analysis: The Case of Expofinder System and Art Exhibitions", Digital *Scholarship in the Humanities,* fqy018, Oxford University Press, 2018). One of the devices developed within the Exhibitium Project is Pathfinder (https://hdplus.es/pathfinder/). Pathfinder is a system for exploring, analyzing and visualizing data about art exhibitions previously recorded in the Expofinder system (www.expofinder.es), a multi-relational database that is semantically enriched with fine-grained metadata. This design is based on the main theoretical framework that underlies the Exhibitium Project: network and complex system theories. At the same time it also fits in with our concept of the art-exhibition domain as an ecosystem (or cultural network structure) built through the different types of relationships that heterogeneous actors (institutions, artists, curators, collectors, critics, etc.) dynamically establish with one another. Figure 1 shows the Expofinder conceptual model represented in a graph.

In general terms, Pathfinder was designed following four essential principles:

1. Flexibility and high precision in the search function. Together with the traditional search functionalities, Pathfinder includes a powerful filter system based on combinations of multiple queries that makes it possible to gather specific subsets of data using as many conditions as needed. Complementarily, the Expofinder multi-relational data model on which Pathfinder operates exponentially increases both the direct and indirect connections established among data. For example, we could retrieve a list of exhibitions held in Madrid, between 2010 and 2015, curated by non-Spanish curators, funded by private entities, where 25–35-year-old female artists participated with pictorial works. Once the search has been run, the list is displayed on the

---

*Speaker

screen with the option of expanding the complete metadata information associated to each exhibition.

2. Optimization of analytical engines for knowledge extraction based on quantitative processes by exploring statistical concepts potentially significant for humanistic research. For this reason, advanced descriptive statistics functionalities have been implemented, thereby making it possible to incorporate previously unexplored categories in the analysis of the art exhibition ecosystems, such as those of 'entropy' and 'outlier'. These categories help us in discovering atypical results that do not respond to the dynamics usually observed. These values are usually an inconvenience when analyzing large data series, but they are an opportunity for the humanist researcher, because they allow us to detect elements outside the limits of 'normality', those that are not exactly the most homogenous with the rest of the group, but rather the ones that stand out among the others (fig. 2).

3. Visualizations as crucial hermeneutic and interpretative tools. According to this idea, a broad array of visualization tools that can be configured by users have also been implemented in Pathfinder with the aim of making the information as insightful as possible. Together with the traditional maps, histograms and networks, others have been designed ad hoc for Pathfinder, such as those that we have called 'geograms' and 'taxograms' (fig. 3). The geoanalytical section is made up of maps which, in addition to the usual georeferenced information and flow paths between connected actors, include a convex envelope calculation based on Delaunay triangulation and Thiessen polygons, which helps us to detect possible areas of influence based on the proximity between different groups of elements.

4. Wide-range usability and interoperability. Finally, Pathfinder can operate with any dataset –whether extracted from Expofinder or not – that matches up with the JSON structure available at https://github.com/antoniocruces/pathfinder. This means that Pathfinder can be used by any researcher or research group interested in making use of the data exploration, analysis and visualization opportunities that Pathfinder offers. At the same time, any subdataset obtained with Pathfinder can be exported in standard formats to be processed using other platforms and software.

Keywords: data infrastructures, art exhibitions, data analysis and visualization, digital art history

Crowdsourcing, scholarly editing and TEI

Wioletta A. Miskiewicz * [1], Pierre Willaime * [1]

[1] CNRS/Archives Henri Poincaré Nancy – CNRS : UMR7117 – France

As a digital editing of the sources, digital humanities call the *natural* materiality of its sources (archives, critical editions, collections...) and consequently their *objectivity* into question. In recent years, owing to proliferation of fakes and hoaxes, the crowdsourcing editing of the humanities archives revive the question of the scientific legitimacy of the *digital scholarly editions*.
In what way can we grant to non-scientists the access to scientific databases without threatening the projects ? To what extent can we allow enlightened amateurs and onlookers to participate to their construction ? And what are the reasons to do that ?

Being convinced that in the new digital ecosystem of science, the scientific legitimacy can be justified only by the construction of enhanced databases, we propose to organize a meeting on the question of the potential that the TEI semantical encoding represents in that context.

We are mainly going to refer to two projects (both projects work in progress): Henri Poincaré's letters (http://e-hp.ahp-numerique.fr, and the Archives éLV (http://www.elv-akt.net/index.php?&langue=en). It's from that last project that we'd like to develop, in an eventual collaboration with another DARIAH project, a new platform of crowdsourcing online transcription. The platform which will be created for the needs of transcription of a philosophical archive (typescripts, manuscripts of research and lectures, letters and diary) faces increasingly the problems tied to multilingualism and therefore with the question of access to the content in the rare languages archives.

We'd like to give some special attention to the project *Testaments de Poilus* (Testaments of the French troops of WWI) https://testaments-de-poilus.huma-num.fr/#!/ directed by E. de Champs (University of Cergy) and F. Clavaud (National Archives) in order to provide experience feedback. This project, whose platform went online in January 2018, allows registered users who have followed a tutorial, the transcription of testaments of the French soldiers died during WWI/The Great War into XML/TEI tags. Emanuelle de Champs will be our guest during this meeting.

Collaborative transcribing is difficult and challenging. On one side, the transcribing interface must be user friendly to allow a massive participation. On the other side, standards of the semantic web such as RDF should be respected to ensure further scientific exploitation.

There is a gap between these two aspects. Projects have made different choices to fill it. The

---

*Speaker

Bentham project (http://transcribe-bentham.ucl.ac.uk/) uses MediaWiki and has developed few extensions for TEI markup support. Testament de Poilus (https://testaments-de-poilus.huma-num.fr/) uses Ace Editor library as a frontend for XML TEI. The Consortium "Archives des ethnologues" offers a different approach with http://transcrire.huma-num.fr/. They use Omeka Classic with scripto's plugin, an integration of MediaWiki adapted for transcripts. This solution without TEI takes advantage of Omeka Classic interface and allows users to indicate metadata for each document.

In the same spirit, The Archives Henri Poincaré (Nancy, France) uses Omeka S to enter and display metadata about Poincaré's letters (http://e-hp.ahp-numerique.fr/). Transcripts of the letters are available in html format and hyperlinks are used to enhance the reader experience. The documentation is properly done in RDF inside the metadata while html transcripts use hyperlinks to refer to the metadata.

None of these solutions is perfect. The power of TEI encoding seems to be hard to conciliate with a user-friendly and simple interface needed for crowdsourcing. Omeka Classic or S offers a intuitive interface to write non-TEI transcripts with the help of Scripto. But there is no link between metadata and transcripts (except with manual hyperlinks which is not a very efficient and standard aware solution).

We'd also like to propose to the Dariah community the question of the usage of TEI encoding in crowdsourcing projects linked to arts and humanities archives, in particular the multilingual ones. Based on the French experience we would like, even before the Warsaw meeting, to make a first round of the question within the forum of DARIAH international community.

PROgressive VIsual DEcision-Making in Digital Humanities: Towards Uncertainty Annotation in TEI

Michal Kozak * [1], Jennifer Edmond [2], Roberto Theron [3]

[1] Poznan Supercomputing and Networking Center – Poland
[2] Trinity College Dublin – Ireland
[3] Universidad de Salamanca – Spain

*** Prepared for submission of both a paper and a poster for the Marketplace ***
Authors: Alejandro Benito (abenito@usal.es), Michelle Doran (doranm1@tcd.ie), Jennifer Edmond (jennifer.edmond@tcd.ie), Michal Kozak (mkozak@man.poznan.pl), Cezary Mazurek (mazurek@man.poz Alejandro Rodríguez (jancho@usal.es), Roberto Therón (theron@usal.es).

In recent years, Digital Humanities (DH) as a research field has experienced a great transformation that has permitted the undertaking of academic projects of great scope and impact, while allowing their immediate exposure to society. At the same time, a number of new and powerful information and communication technologies (ICTs) have made possible the exploitation of a wealth of data (either digitised or born digital) that have changed enormously the practice in DH. From the creation to the consumption of digital resources, there are new stakeholders, contexts, and tasks to consider. The volume of digital resources produced (or digitised), stored, explored, and analysed in any DH project is immense. Therefore, traditional humanities tools for managing uncertainty have to be either substituted or aided with ancillary tools in the form of interactive visualisations or novel user interfaces in this environment. Furthermore, during the whole lifecycle of any DH project - from the data preparation to the actual analysis or exploration phase - many decisions have to be made in order to yield the desired results that depend on managing the uncertainty pertaining to both the datasets and the models behind them and which may in turn introduce their own uncertainty into the research process.

This presentation will introduce the ongoing work of the interdisciplinary PROgressive VIsual DEcision-Making in Digital Humanities (PROVIDEDH) project, a three-year project funded within the CHIST-ERA call 2016 for the topic "Visual Analytics for Decision Making under Uncertainty – VADMU." The project aims to give DH scholars a space to explore and assess the completeness, evolution, and interconnectedness of digital research objects, the degree of uncertainty that the models applied to the data incorporate, tolerate or introduce, and to share their perspectives and insights with the project's broad range of stakeholders.

The project itself can be broken down into the following research objectives:

To establish a taxonomy of sources of uncertainty that may appear during the lifecycle of a DH project.

---

*Speaker

To develop a set of metrics that convey the degree of uncertainty that research objects, data sets, and collections introduce as well as the different computational models applied to them.

To propose a framework that makes use of the uncertainty metrics, so any given representation of the data can be assessed according to its degree of uncertainty.

To propose a Progressive Visual Analytics solution that ensures that users are able to trace changes in data and its inherent uncertainty as well as in the way it is perceived.

To develop a web-based multimodal collaborative platform for the progressive visual analysis of different DH collections, both for scholars and citizen scientists.

To trigger the formation of a "community of practice" that humanists can build on to reinforce each other's efforts to achieve metrics that are both practical and of high quality.

The objective of the presentation is to share information and resources relating to the use of the PROVIDEDH collaborative platform. In addition to enabling progressive visual analyses of large research datasets, users of the platform will be able to perform a close reading of annotated TEI P5 files with a focus on uncertainty. Encoders of text often find it useful to indicate that some aspects of the text are problematic or uncertain, and to indicate who is responsible for various aspects of the markup. A standard way to annotate DH research datasets is through the use of TEI tags. Although the TEI provides various methods to indicate that some aspects of the encoded text are problematic or uncertain, it is not a common practice and uncertainty remains a challenge that any researcher using the dataset has to face. The PROVIDEDH project intends to bring uncertainty to the surface by associating the existing TEI elements with the developed taxonomy of sources of uncertainty and by implementing an environment for progressively visualising uncertainty in humanities data.

The talk will focus on visualisations and the user-friendly interface of annotating electronic text implemented in the close readingmodule of our collaborative platform. Solutions for many possible cases will be presented, like annotating the same or intersecting texts by many people with different profiles, or annotating already tagged entities by many people with different perspectives of those entities. Furthermore, potential directions of tracking and processing these annotations will be outlined.
As an ancillary objective of the presentation, it is hoped that it will encourage the attendees of the annual event to join the project's interdisciplinary and international "community of practice". The presentation will be accompanied by a poster for the Marketplace.

Keywords: Progressive Visual Analytics, Uncertainty, Taxonomy of Uncertainty, TEI, Humanities Research Data, Annotation.

The Base of Knowledge in the field of humanities - a new tool to share data and information and to support research management

Jakub Koperwas * [1], Łukasz Kobyliński [2], Jolanta Stepniak * [3]

[1] Politechnika Warszawska – Poland
[2] Instytut Podstaw Informatyki – Poland
[3] Sages Sp.z o.o. – Poland

The aim of the paper is to point the main directions of IT development to support the policy of open science (open research data) and open access in humanities. Omega-Psir is an open source tool generating the base of knowledge for effective management of research institutions (universities), dissemination of information and data, and for purposes of cooperation and sharing knowledge with others. Omega-Psir is the tool to create three kinds of systems in one: an institutional repository, the researchers profiling system and the system of researchers assessment and reporting for management (including reporting for governmental units).

The unique idea of the system is to create and maintenance the Base of knowledge, with wide scope of data (publications, PhD thesis, project reports, achievements in professional activities - such as participation in editorial boards, conference committees, being spokesman or expert, author of technology or piece of art, project leader, etc.). All those data, together with the full text of publications, metadata of research data, external metrics and ontologies are used to create automatically researchers and units specialization profile. The profiling system is presented in visual form to improve the perception of information and by doing so acquire a better understanding of the profiles and area of expertise. User-friendly sharing documents and data functionalities offer additional benefits to increasing numbers of citations, and easier way to find teams and specialist for further cooperation.

The system is compliant with OAI-PMH standards, so it could be indexed by the worldwide known archives of open access resources (ex. DOAR, DOAB, OAIster, etc.).The system communicates with the WoS and Scopus, as well as with Google Scholar - to import a number of citations and metrics (IF, Scopus CiteScore, Scopus SJR), as well as for other data import. Communication to Sherpa-Romeo database provides information about the license policy of journals and publishers. Omega-Psir is also guaranteeing the author right to determine the policy of openness and type of license for shared data and documents.

The research institutions with a humanistic profile, which implemented Omega-Psi software significantly improved visibility of achievements and cooperation with other researchers. Omega-

---

*Speaker

Psir, Polish software implemented in over 20 research institutions (including University of Gdańsk and University of Opole) is still developing tools to offer broad access to research data, an important element of open science development.

Developing symbolic and audio data analysis modules for a prototype traditional music repository

Ewa Lukasik * 1, Ewa Kuśmierek * 2

Ewa Dahlig-Turek * 3

Jacek Jackowski 3, Magdalena Chudy 3, Marcin Werla 2, Tomasz Parkola 2

1 Institute of Computing Science, Poznan University of Technology – Poland
2 Poznan Supercomputing and Networking Center – Poland
3 Institute of Art, Polish Academy of Sciences – Poland

Music Information Retrieval Working Group has been officially established by DARIAH-PL Consortium Board in March 2018. The group's objective is to develop and implement state-of-the-art music information retrieval (MIR) tools for large-scale interdisciplinary research in musicology and ethnomusicology. In particular, for scholars working on various aspects of traditional music, a virtual research environment (VRE) is being built based on dLibra Digital Library Framework that has been developed by Poznan Supercomputing and Networking Center (PSNC) since 1999 [1]. The system provides an extended support on: i) data collection and management; ii) data processing, analysis and visualisation; and iii) result publishing.

The sound archive, a selected part of which is to be incorporated in the VRE, is currently administered and operated using a dedicated system designed and built in the Institute of Art of Polish Academy of Science (IAPAS) to address its specific functional requirements [2]. The system includes a dedicated database with a structure optimized for developing ethno-phonographic collections and a specialized user application. The database stores descriptive, administrative and technical metadata which describes each digitized medium, its creators and performers, the object's structure, the date of object creation, technical data related to the object, paradata regarding its digitization process, etc.

The works on the VRE started with creation of a prototype MIR content repository based on dLibra digital library software with a sample of IAPAS phonographic collections. Those phonograpic sample records contain metadata describing analog media and their content, audio files, scans of archival documents including song lyrics and related photographs. The objects are organized into several thematic collections. Using the dLibra package, the repository has been set up in a relatively short time required to configure a new library instance tailored for specific MIR needs related to metadata schema, music object characteristics, collection structure, and users access rights.

---

*Speaker

As a part of VRE framework, two modules were built in their prototype form: 1) a web-based application (WebEsAC) for archiving and retrieval of Polish traditional music in the symbolic text-based EsAC format; 2) a web-based application (SoundScribe) for automatic melodic transcription and correction of traditional music from audio to symbolic format (MIDI and others).

WebEsAC module offers an online access to a database containing symbolic representation of tunes, mainly acquired from Oskar Kolberg (1814-1890) collection preserved in music notation, where each tune is saved as a separate record with a structure and melody representation in EsAC format. EsAC was developed for monophonic music, with European folk song databases in mind. The implementation of EsAC-to-MIDI converter gives additional flexibility to the system and allows aural verification of encoded melody and rhythm. There were already two software releases of WebEsAC. The second release extended the initial functionality of the module. First of all, user access and rights management was added to prevent the unauthorized access to the repository, e.g. user registration, login, authorization, email verification, and password change. Further enhancements include mechanisms for preventing syntax errors (e.g. validation of the EsAC records when adding single or multiple records in txt format to the repository), extracting the scale from each tune, converting melodic line into a sequence of intervals, calculating statistics of scale-types and rhythmic bar structures, multicriteria retrieval of tunes from the database. Further functions, including pitch and rhythm extraction, and MIDI-to-EsAC conversion, are planned to be developed.

The module for automatic melodic transcription and correction is based on pYIN algorithm [3]. Available as a Vamp plugin, it was adapted as a dll library and used for extracting the predominant melody of an audio recording and converting it into MIDI data and music notation that is saved in an XML file. Additionally, a txt file is generated, containing detailed information about the fundamental frequency value sampled every 0.01 ms. The MIDI and XML outputs together with the source audio file are uploaded to dLibra using editor application and are published in a dedicated dLibra folder format. It is possible to load a single audio file or a folder containing many files or subsequent folders with files. A front end application was also built to enable correction and edition of automatically generated MIDI data in a visual form. This module has positively passed user tests and will be further improved.

Future works will focus on improving the new dLibra music modules and equipping them with new features that allow further-reaching comparative work in the field of traditional music.

M. Werla et al. (2018) "Developing Music Digital Library based on Polish Traditional Music Archives and dLibra", in Proceedings of the 5th International Conference on Digital Libraries for Musicology, Paris.

J. Jackowski (Ed.). (2017) "Polish traditional music – phonographic heritage. The current state – preserving and sharing" (in Polish). Liber Pro Arte, Warsaw.

M. Mauch and S. Dixon, (2014) "pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Florence.

Author Index