

DARIAH Annual Event 2020: Scholarly Primitives

Book of abstracts

Programme Committee

Toma Tasovac (Chair)
Femmy Admiraal (Co-Chair)
Francesca Morselli (Co-Chair)
Koraljka Kuzman Šlogar
Agiatis Benardou
Sally Chambers
Marianne Huang
Andrea Scharnhorst
Susan Schreibman
Andreas Witt

Table of contents

WG meeting	6
DH Course Registry WG Meeting: Between Reaching Out and Future Technical Developments, Tanja Wissik [et al.].....	7
DARIAH ELDAH working group meeting, Vanessa Hanneschläger [et al.]	9
Bibliographical Data Working Group Meeting, Vojtěch Malínek [et al.]	11
Research Data Management Working Group meeting, Erzsébet Tóth-Czifra	13
WG Meeting GiST (Guidelines and Standards): Documenting standards - sustainable approaches, Klaus Illmayer [et al.].....	15
#dariahTeach: Moving Forward. Listening to the Community, Susan Schreiberman [et al.].....	17
Sustainable Publishing of Metadata: Open Data, Johan Van Der Eycken [et al.].....	19
Workshop	22
Cracking digital archival research and metadata: Archives Portal Europe, Marta Musso [et al.].....	23
Building trustworthy repositories: Introduction to CoreTrustSeal certification, Tomasz Parkola [et al.]	25
Addressing legal and ethical dimensions of the processing of personal data in research contexts: The ELDAH Consent Wizard as a tool for fostering best practices for consent and transparency in DH., Walter Scholger [et al.].....	27
Using Collaborative Annotation to Explore and Analyse Uncertainties in Digital Humanities Datasets, Michelle Doran [et al.].....	29

Does Communication belong to Scholarly Primitives? Innovations in scholarly communication through the lenses of Digital Humanities, Erzsébet Tóth-Czifra [et al.]	31
Bibliographical Data Workflows: discover, analyse, and improve them, Vojtěch Malínek [et al.].....	33
Information Extraction Workflow for Digitised Entry-based Documents, Mohamed Khemakhem [et al.].....	36
Poster/demo session	39
CLARIAH-DE - Aligning two research infrastructures: Experiences and Challenges, Stefan Buddenbohm [et al.].....	40
CLARIAH-DE - Added value of the infrastructure merger, Stefan Buddenbohm [et al.]	42
The SSH Open Marketplace: How to contribute to this discovery portal, Stefan Buddenbohm [et al.].....	44
The DARIAH ELDAH consent form wizard, Vanessa Hanneschläger [et al.]	46
Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism - The State of the Art, Sanja Peric Gavrancic [et al.]	48
TaDiRAH as Linked Open Data, Canan Hastik [et al.].....	50
Digital scholarly primitives : how data curation/implementation/aggregation and mapping happen in ArkeoGIS, Loup Bernard	52
Education as a Scholarly Primitive - The DH Course Registry, Hendrik Schmeer [et al.]	54
The database of antique archaeological sites of the Republic of Croatia, Kristina Turkalj [et al.].....	55
Supporting Decision-Making in the Entity Normalization Task. On the Example of PROVIDEDH., Grzegorz Bryk [et al.].....	56
DARIAH contributions - visual analytics and workflows, Andrea Scharnhorst [et al.]	58
Communicating: the forgotten, yet indispensable scholarly primitive and its infrastructural dimension, Maciej Maryl [et al.]	59

BBT User stories. Benefits from joining the thesaurus federation, Helen Goulis [et al.]	61
Integrating European Literary Bibliographical Data: Open Bibliodata Initiative, Vojtěch Malínek [et al.]	63
DARIAH-EU "Bibliographical Data" Working Group 2020, Vojtěch Malínek [et al.]	66
Introducing the AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) multilingual research tool, Róbert Péter [et al.]	68
Project Znameniti: spatial humanities and deep mapping, Nevenka Kalebic [et al.]	70
Insights on scholarly primitives from Digital Humanities research in Spain, Maurizio Toscano [et al.]	72
OPERAS services - Metrics Service, Certification Service, Publication Service Portal and Discoverability Service, Yoann Moranville	74
Echo from the War: Preserving the Memory of the Bulgarian Veterans from WWII, Konstantin Golev	76
Open Badge Ecosystem for the Recognition of skills in Research Data management and sharing, Tomasz Umerle	78
Retroconversion of bibliographical publications: main stakes and challenges, Patryk Hubar [et al.]	80
Towards a Global Study of Algorithms thanks to Digital Humanities and Sociology, Jeremie Poiroux	82
Digging for unspecified information requirements: a case study of Digital Library of Arne Novák users, Michal Lorenz [et al.]	84
DARIAH public consultation	85
The SSH Open Marketplace – Involvement of the research community in the development process, Laure Barbot [et al.]	86
Scholarly Primitives in Theory and Practice	88
Digital Humanities or humanities in digital?, André Pacheco	89
Scholarly primitives from a workflow perspective, Marijn Koolen [et al.]	91

Beyond scholarly primitives: an epistemological foundations for computational research in the humanities, Fabio Ciotti	93
Reassessing Scholarly Primitives through Experience from Project Factories and Urban Fabric in Bulgaria, Ivaylo Nachev.....	95
Scholarly primitives in the field of Glagolitic script: foundations for transition of traditional research activities into virtual research environment (VRE) for Zadar Glagolitics, Marijana Tomić [et al.]	96
3D Scholarly Editions: Scholarly Primitives Reboot, Susan Schreibman [et al.] ...	98
Tools and Workflows	100
Unus pro omnibus! Generic research tool for all Humanities disciplines., André Kilchenmann [et al.].....	101
Making the most of digital resources with IIF and OCR enabled transcription toolset, Tomasz Parkola [et al.].....	103
Criteria for Recording and Categorizing Scholarly Digital Editions, Simon Sendler [et al.]	105
The CITE architecture – assigning knowledge to locations in the text, Neven Jovanovic [et al.].....	107
Approaches to Collaborative Multilingual Thesauri Development in Educational Context, Goran Zlodi [et al.]	109
Modelling the socio-linguistic processes using the NLP tools for syntactic parsing, Neo4j Graph database for storing and semantic and communication analysis, Benedikt Perak.....	111
Mediating the Past	113
The Past is a Foreign Country: the Digital Rediscovery of the 19-c. Documents from the Metropolitan See of Philippopolis, Dimitar Iliev	114
Integrating archival materials for the study of the turbulent Greek 40s, Vicky Dritsou [et al.].....	116
e-Version of the Republic of Letters, Sepideh Alassi [et al.].....	118
Infrastructural Challenges	120

CLARIAH-DE - Aligning two research infrastructures: Experiences and Challenges, Stefan Buddenbohm [et al.]	121
Services in a research infrastructure - an attempt to review current definitions and their implications, Laure Barbot [et al.]	123
Digital libraries as part of the digital humanities infrastructure: services, tools and programmes, Renata Petrušić [et al.]	125
Synergy session	126
Geohumanities Data: Managing, Using, Opening, Sharing, James Smith [et al.]	127
DH process graphs, Panos Constantopoulos [et al.]	129
The value of collaboration. Challenges and opportunities for the Arts, Humanities and Social Sciences, Francesca Morselli [et al.]	131

WG meeting

DH Course Registry WG Meeting: Between Reaching Out and Future Technical Developments

Tanja Wissik *¹, Hendrik Schmeer *²

¹ Austrian Center for Digital Humanities (ÖAW) – Austria

² CLARIN ERIC – Netherlands

Designed as a hub to collect information on DH teaching activities across Europe and beyond and to increase the visibility of these teaching activities beyond the usual university networks, the DH Course Registry (DHCR) is a search platform that allows users to access metadata about DH teaching activities. In the last years, we have made a lot of progress in terms of dissemination, e.g. holding the DH Course Registry Metadatathon, printing and distributing postcards etc., and in terms of technical development, e.g. API development. However, we encountered different challenges in these two areas, which we would like to discuss with you.

- **Reaching out:** Even though there were many dissemination activities going on centrally as well as on a national level regarding the DH Course Registry, there is a lack of a common dissemination strategy for it. The special challenge in the case of the DH Course Registry is that the registry does not have its own dissemination channels, except for the platform itself. Since it is too time-consuming to create new channels and make them known, it is more effective to create a strategy which uses the channels already established by CLARIN and DARIAH. Additionally, we would also consider and ultimately plan the use of the research infrastructures, their national nodes, the channels of the contributors etc. in an efficient and resourceful way.

In the WG Meeting, we would like to discuss these topics and explore how national CLARIN and DARIAH nodes and National Moderators can be involved more actively in dissemination campaigns. Furthermore, we would like to collect new ideas and explore how initiatives like the Call for contributions <https://dighumlab.org/dhcr/> of the DIGHUM LAB Denmark can be used as an example for other countries.[1]

-
- **Technical Development Roadmap:**

*Speaker

Since the creation of the new front-end and API, the code structure has dramatically changed and the code quality was improved by a partial refactoring. Up to now, only maintenance of the platform is provided by CLARIN, while further development depends on limited project funding. For future platform development and continuation of the service, more actions need to be taken to improve code quality and maintainability.

In the meeting, we want to explore possibilities for contributions or involvement of additional personnel resources on the technical side. The envisioned dissemination strategies will require constant development of new features, further UI improvements and test coverage to keep up with growing demands and provide stable service.

Besides members of the WG (e.g. National Moderators), we also welcome everybody to the meeting who wants to contribute to the future development of the DH Course Registry, especially dissemination and technical experts.

This work will be supported by the DARIAH WG Funding Call 2019/2020 in the "Reaching out: Dissemination Strategy and Planning for the DH Course Registry" Project <https://www.dariah.eu/2019/11/25/working-groups-funding-scheme-2019-2020-meet-the-winning-projects/>

Keywords: Registry, Education, Outreach, Technical Development

DARIAH ELDAH working group meeting

Vanessa Hanneschläger ^{*} ¹, Walter Scholger ^{*†} ², Koraljka Kuzman
Šlogar ^{*#} ³

¹ Österreichische Akademie der Wissenschaften – Austria

² University of Graz (Uni Graz) – 8010, Universitätspl. 3, 8010 Graz, Austria

³ Institute of Ethnology and Folklore Research (IEF) – Šubićeva 42, 10000 Zagreb, Croatia

The DARIAH Working Group ELDAH (Ethics and Legality in Digital Arts and Humanities) will meet for a working group meeting at the DARIAH annual event 2020 in Zagreb. Anyone is welcome to join this meeting to learn about the group and its activities; interested parties can also take a look at the ELDAH blog (<https://eldah.hypotheses.org/>) to get an overview of the group's many activities.

At this meeting, the group would like to discuss the progress of the Consent Form Wizard project (CFW), plans for organizing a discussion forum at the DH2020 conference in Ottawa on the topic of "Global Copyright Laws, Knowledge Protection, and the Impact on Open Digital Humanities", and other workshop and outreach activities in collaboration with other infrastructures and projects such as the CLARIN Legal Issues Committee (CLIC) and the SSHOC project.

The meeting will be organized by the working group co-chairs Koraljka Kuzman Šlogar (DARIAH-HR), Walter Scholger (CLARIAH-AT) and Vanessa Hanneschläger (CLARIAH-AT). We expect to have 25 to 30 participants for this meeting of a very active working group.

Preliminary agenda:

Welcome by the WG chairs

Introduction round

Presentation of Consent Form Wizard project (CFW)

Collection of questions and input for the discussion forum at the DH2020

Presentation of further planned activities in 2020 (workshops, webinars)

Brainstorming: Collection of ideas for activities in 2021

*Speaker

†Corresponding author: walter.scholger@uni-graz.at

#Corresponding author: koraljka@ief.hr

Keywords: legal issues, consent form wizard, copyright, data privacy

Bibliographical Data Working Group Meeting

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

Main Organizers:

Vojtěch Malínek

Tomasz Umerle

Main programme points:

Summary of WG's activities in recent months, and its tasks for 2020. Discussion about the current state of WG's report on bibliodata in the humanities.

Presentation of individual members' bibliodata projects.

AOB.

Preliminary agenda:

First part (cca 30 minutes) of the meeting will be devoted to brief summary of WG's activities since its establishment in the end of 2019, and tasks for 2020, introduction of new members, and short discussion around internal workflow of the WG, and its modes of communication.

Second part (cca 60 minutes) will be devoted to presentation of the current state of the main WG's task for 2020, which is the preparation of the report "Analysis of Bibliographical Data Landscape: Bibliodata curation, research, and collaboration in the humanities". Presentation will include case studies by WG members which will highlight the report's main findings, and it will outline the next steps leading to report's publication.

We welcome all those who are interested in our WG's agenda, also - in learning about the

*Speaker

†Corresponding author:malinek@ucl.cas.cz

report, and contributing to it.

Envisioned participants (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

Matteo Romanello, Digital Humanities Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland

David Lindemann, University of Ljubljana, Slovenia

Maciej Maryl, Polish Literary Bibliography, Institute for Literary Research of the Polish Academy of Sciences, Poland

Also, if technically possible, there will be multiple members attending on-line, in case they could not come in person (which is not clear right now).

Keywords: bibliographical data, working group meetings

Research Data Management Working Group meeting

Erzsébet Tóth-Czifra * ¹

¹ Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

Although at the time of writing this proposal, the Research Data Management Working Group is still in the final phase of its accreditation process, we would love to take advantage of the DARIAH event to organize our first ever face to face meeting as well as to meet potential new members. Since the topic of research data management is a research support area that is relevant to all scholarly fields, our WG aims to build on and relate the work of several DARIAH WGs. To this end, our participation and contribution of the Annual Event would give us a unique opportunity to continue our efforts to find the best possible ways to build on and help their work will be among the first activities of the proposed WG.

Main topics of the meeting: 1. Introduction of the nascent WG and its aims to potential new members 2. Making progress on our work plan for 2020.

In the first part of the meeting, we would like to introduce our group, its aims and scope and the work done and to be done to the interested broader community. To engage them, we would also give them the opportunity to briefly present the type of data they are working with together with the associated data management challenges they are facing in their scholarly workflows. In the second part of our meeting, we aim to form smaller groups and targetedly work on our 4 main task forces:

- Building a knowledge base of data management best practices in the humanities (Zotero library, discussing possibilities to its further enrichment such as annotations, community tagging etc.)
- Providing concise, discipline-or data type-focused model workflows with the guidance of disciplinary champions (4 subtask leaders are confirmed so far)
- Supporting the exchange between the GLAM sector and scholars under the aegis of the Heritage Data Reuse Charter
- Developing criteria for good data management plans and model DMPs in the humanities domain.

The meeting will conclude with a joint reporting session and discussion the next steps if the establishment of the WG. As a further step in the direction of working closely with other DARIAH

*Speaker

WGs, we also submitted a joint synergy session proposal with the Geohumanities WG.

Envisioned Participants (some will be present online, for example via Skype):

- Erzsébet Tóth-Czifra (WG chair, DARIAH EU)
- Laurent Romary (INRIA, France)
- Erik Buelinckx (Royal Institute for Cultural Heritage, Brussels)
- Johan Van der Eycken (State Archives of Belgium)
- Sally Chambers (University of Ghent)
- Ulrike Wuttke (FH Potsdam, Germany)
- Rene van Horik (DANS, The Netherlands)
- James L. Smith (University College Cork)
- Marta Blaszczyńska (Digital Humanities Centre, Institute for Literary Research of the Polish Academy of Sciences, Poland)
- Maciej Maryl (Digital Humanities Centre, Institute for Literary Research of the Polish Academy of Sciences, Poland)
- Jasenka Ferber Bogdan (The Fine Arts Archives, Croatian Academy of Sciences and Arts, Croatia)
- Unmil Karadkar (University of Graz)
- Péter Király (GWDG, Germany) - could only attend if the meeting would not overlap with the Bibliographical data WG meeting
- Tomasz Umerle (Polish Academy of Sciences) - could only attend if the meeting would not overlap with the Bibliographical data WG meeting
- Walter Scholger (University of Graz) - could only attend if the meeting would not overlap with the ELDAH WG meeting

Keywords: research data management, data sharing, long, term access, cultural heritage, cultural heritage data reuse charter, data management plans

WG Meeting GiST (Guidelines and Standards): Documenting standards - sustainable approaches

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 ENPC – École des Ponts ParisTech (ENPC) – France

Main organisers:

Klaus Illmayer
Marie Puren
Charles Riondet

Brief description of the main aim:

Based on the working plan of the WG GiST (Guidelines and Standards) and the experiences made last year, this meeting of the WG is under the topic of sustainability. There are three projects for this year, that are in the center of the activities of the WG. In the WG meeting we will present them, explain how they are connected together, and discuss how they contribute to the establishment of a sustainable ecosystem for the WG.

(1) Handover of the SSK and creating a long-term perspective.

The SSK (Standardization Survival Kit: <http://ssk.huma-num.fr>) was developed in the Horizon 2020 funded project PARTHENOS. This project ended last year, therefore the WG is taking over the responsibility for the SSK. We are discussing how to feed new content and maintain the available scenarios. There will be a focus on keeping up-to-date scenarios as well as assessment methods.

*Speaker

†Corresponding author:klaus.illmayer@oeaw.ac.at

(2) Development of a strategy for watching unattended and emerging standards.

An important aspect of the WG is to watch out for current trends in standards. Therefore implementing a strategy to fulfill this task is an important focus of the WG in 2020. There are different approaches possible: establishment of a standards blog, extension of the SSK, building up a standards database. We will discuss these approaches and decide on a strategy.

(3) Create a module for #dariahTeach on standards.

The primary task for 2020 will be the creation of a teaching module on standard awareness for #dariahTeach. It should also give insights into the elaboration of scenarios for the SSK. We will discuss the structure of the teaching module and will do a brainstorming for the teaching units.

In the wrap-up part of the WG meeting we distribute the work that is still to be done for the three presented activities. We would be very happy if there are new members for the WG that decide to join us.

—

Agenda (2-hour slot):

10 min: Introduction, presentation of the main aim of the meeting

30 min: SSK

20 min: Strategy for collecting emerging standards

50 min: Module in #dariahTeach

10 min: Wrap-up and distribution of tasks

—

List of envisioned participants:

Main organisers

Members of the WG

Interested participants

#dariahTeach team members (TBR)

Keywords: standards, guidelines, sustainability, SSK, dariahTeach

#dariahTeach: Moving Forward. Listening to the Community

Susan Schreibman ^{*} ¹, Costas Papadopoulos ^{*} [†]

¹ Maastricht University – Netherlands

#dariahTeach is an asynchronous teaching and learning platform designed as open-source, multilingual, community-driven resource for high-quality teaching and training materials for the digital arts and humanities. The platform currently has six courses and two workshops developed in English (three of them translated in French, Hungarian, and Greek; Spanish versions are also under development). Its 80 videos on YouTube have been viewed over 50.000 times. In April 2018 several members of the original #dariahTeach team were awarded a grant for IGNITE: Design Thinking & Making in the Arts & Sciences which builds on #dariahTeach to develop a 20 ECTS module that fosters cross-sectorial curricula by combining entrepreneurship and technology with arts and heritage. The Geohumanities Working Group is also developing learning materials.

#dariahTeach is currently transitioning into a new phase, harnessing and channeling previous investment by offering a one-stop-shop for peer-reviewed, quality assessed publication of training and teaching materials delivered within a homogenous framework. This will enable the long-term growth and development of #dariahTeach towards the establishment of a novel organisational model for sustaining peer-reviewed open access teaching materials. We believe that training materials deserve to be recognized as publications in their own right, and therefore, the main goal for the next phase of #dariahTeach is to get recognised as a publication, functioning like any other peer-reviewed, community-driven journal. We also aim for raising the peer-review model that #dariahTeach will establish at the forefront of new innovative models of publishing peer-reviewed, collaborative creation. #dariahTeach was recently awarded a sustainability funding from DARIAH that would enable this transition. This grant facilitated the organisation of a #dariahTeach Publishing Board meeting as part of the annual DARIAH event (Warsaw 16-18 May 2019) in which the members of the group discussed the Editorial Structure, and potential publication lifecycle.

The aim of the proposed Working Group meeting is to use the community forum that the DARIAH Annual Event provides to discuss with existing and potential #dariahTeach contributors the best ways to move this project forward. Since the project is mostly based on voluntary work and given the limited set of resources, we would like to gauge the community's reaction to this endeavour before we move forward. How could community members contribute? What resources do they need? How could #dariahTeach teaching and learning content be integrated into their teaching/research practice? What support would they like to see #dariahTeach providing them with in order to be able to contribute to the project, e.g. by developing a teaching unit (course, workshop, training etc.), translating or adding to existing content? What synergies

^{*}Speaker

[†]Corresponding author:k.papadopoulos@maastrichtuniversity.nl

could be developed with other Working Groups and teaching/research projects so this new phase can be socially sustainable, thus ensuring an ongoing, productive relationship between its users and its creators? The goal of this Working Group meeting is to provide us with a more concrete plan for the future of #dariaTeach, while also helping us attract new content, and helping potential collaborators/ contributors make the decision to share and publish their teaching/training materials.

Keywords: education, infrastructure, online learning, eSkills, lifelong learning

Sustainable Publishing of Metadata: Open Data

Johan Van Der Eycken *^{1,2}, Dorien Styven *^{† 3}

¹ KU Leuven – Faculteit Letteren in Leuven Blijde Inkomststraat 21 bus 3301 3000 LEUVEN, Belgium

² Algemeen Rijksarchief – Ruisbroekstraat 2 1000 Brussel, Belgium

³ Kazerne Dossin (KD) – Goswin de Stassartstraat 153, 2800 Mechelen, Belgium

Chair

- Johan Van der Eycken (State Archives)
- Dorien Styven (Kazerne Dossin, EHRI)
- Tom Gheldof (KU Leuven)
- Eric de Ruijter (International Institute of Social History)

Agenda

- Overview of current projects
- Presentation of SODA (Social Science Data Archive)
- Presentation of BISHOPS (Belgian Infrastructure for Social Sciences and Humanities Open Science)
- Optional: Strategies of: Netwerk Digitaal Erfgoed Nederland
- WG Activities
- Evaluation of past activities

- Collaboration with SSK

- Belgium and the open science cloud (state of affairs)

*Speaker

[†]Corresponding author: dorien.styven@kazernedossin.eu

- Future projects: introduction and brainstorm session
- Cross-WG and Infrastructural collaboration
- WG-Communication: brainstorm and task division / assignment

After welcoming the participants the coordinators will first **introduce** the goals and activities of the working group as we have learned at the past annual events that the WG meeting attracts both established members of the WG as well as newcomers interested in the work done by the WG.

Then several **projects represented within the WG** are presented by the project leaders:

(1) A first presentation focuses on the Social Science Data Archive project or SODA. SODA will be finished by the time the annual event takes place which allows the WG to present the project's concrete results and discuss them with the WG participants. The project leader will give a live demonstration of the functional SODA infrastructure.[1]

(2) The second presentation is dedicated to the BISHOPS-project (Belgian Infrastructure for Social Sciences and Humanities Open Science).[2] This project, funded by Belgian Science Policy, builds on the results of SODA. BISHOPS aims to create a platform for storing, archiving and making research data available for reuse as open data. The current state of affairs as well as the results of the running analysis (researchers' expectations, metadata and the legal aspects (GDPR, Licenses) will be presented. Input from the WG participants is much appreciated.

Following both presentations, **the WG's own activities** are analysed:

(1) The finalised and running activities are presented briefly and their results will be discussed by the participants in order to decide on their success and a possible relaunch in the future, e.g. the Trust and Understanding-workshop organised in Brussels in May 2018 or the publication of the contributions in a special edition of the *Archief- en Bibliotheekwezen* journal (A-rating).

(2) The participants will brainstorm about future activities within the scope of the WG.

(3) The participants will discuss how the WG can intensify its collaboration with related Workinggroups (Eg. SSK, DRM-WG) and European research infrastructures such as Archives Portal Europe[3] (APE), European Holocaust Research Infrastructure[4] (EHRI), Distributed System of Scientific Collections[5] and E-HRIS[6]. To the extent possible, members of these infrastructures are invited to participate.

The last topic of the WG meeting will be dedicated to the **WG communication**. This section will consist of a brainstorm session during which the participants will examine how the activities organised by and the expertise represented in the WG can be better publicized and validated, e.g. via blog entries[7] and other media. Concrete tasks regarding communication will be given to WG members in order to boost the WG communication within a set amount of time.

This WG meeting is public and can be attended by all interested parties.

<https://www.cessda.eu/News-Events/News/CESSDA/SODA-A-roadmap-towards-a-Belgian-data-archive-in-2020>; <https://sodabelgianproject.wixsite.com/sodaproject>

<http://www.arch.be/index.php?l=nl&m=lopend-onderzoek&r=onderzoeksprojecten&pr=bishops-belgian-infrastructure-for-social-sciences-and-humanities-open-science>

<https://www.archivesportaleurope.net>

<https://www.ehri-project.eu/>

<https://www.dissco.eu/>

<http://www.e-rihs.eu/>

<https://www.dariah.eu/tools-services/tools-and-services/tools/hypotheses-org-academic-blogs/>

Keywords: Open Data, Research Data, RI

Workshop

Cracking digital archival research and metadata: Archives Portal Europe

Marta Musso ^{*† 1}, Kerstin Arnold ^{*2}, Johan Van Der Eycken ^{*3}, Frank Fischer ^{*4}, Laure Barbot ^{*5}

¹ King's College London – United Kingdom

² Archives Portal Europe – Netherlands

³ DARIAH's WG on Sustainable Publishing of Metadata – Belgium

⁴ DARIAH – Russia

⁵ DARIAH – Germany

Archives Portal Europe (APE) is an aggregator that connects the catalogues and digitised archival material of archives from more than 30 countries in a single research point. It is currently the largest online archival repository in the world, and it aims to connect all archival collections related to the history of European countries, from the large state archives to local and private archives. The portal allows to conduct research with an unprecedented outlook based on cross-country archival comparisons and multilingual search. Thanks to this new tool, the functions indicated by Unsworth as "scholarly primitives" are available to researchers, and expressed in new forms.

This half-a-day workshop will present APE to both researchers and archivists as a tool that allows new types of historical research. Researchers will learn how to conduct research on APE in the most effective way possible; archivists will be presented with the metadata solutions applied by APE in order to provide the best possible returns to the users' search queries, while at the same time being sustainable and feasible for the partner institutions.

The workshop will consist of two parts: in the first part (45' plus 15' break), an overview of the portal and its research features will be presented to the attendees by Dr Marta Musso, Research Manager for APE and Teaching Fellow at King's College London. The presentation will include a few exercises for attendees to experiment first hand with portal search and the guidelines provided on how to better pursue digital research using the portal.

In the second part of the workshop (60'), a more general reflection on how metadata can inform researchers will take place. The metadata policies and politics of APE will be presented by Mr

*Speaker

†Corresponding author:marta.musso@archivesportaleurope.net

Stefan Papp, developer for the portal, who will provide insights in the back-end perspective; Mr Johan Van Der Eycken will present the work of DARIAH's Working Group on Sustainable Standards; Prof Frank Fischer and Laure Barbot will present DARIAH-EU involvement in the Social Sciences and Humanities Open Cloud (SSHOC) project. Dr. Klaus Illmayer (t.b.c.) will make the connections to the Parthenos project. Following these presentations, speakers and attendees will be invited to reflect together on the role of metadata and cloud computing in digital archival research, and how archivists and researchers can collaborate to make digital archives as open as possible by improving searchability and retrievability.

Requirements:

4-hour workshop

Projector and good wifi connection

Bring Your Own Laptop

Target audience

The workshop is mainly aimed at history researchers, from any period and topic, who include strong elements of archival research and work on primary sources, particularly in a multi-national comparative perspective.

However, the workshop is also open to any archivist who wants to collaborate with archives goers and researchers in order to reflect on how to improve the functions of retrievability and searchability in a digital environment.

Deliverables

A document holding the output of the conversation, to become a series of guidelines / best practice for metadata and research.

Keywords: metadata, Archives Portal Europe, digital archives, digital history, scholarly primitives, SSHOC

Building trustworthy repositories: Introduction to CoreTrustSeal certification

Tomasz Parko-la ^{*† 1}, Rene Van Horik ^{*‡ 2}, Birger Jerlehag ^{*3}, Daan Broeder ^{*4}

¹ Poznan Supercomputing and Networking Center – Poland

² Data Archiving and Networked Services – Netherlands

³ Swedish National Data Service – Sweden

⁴ KNAW Humanities Cluster/CLARIN ERIC – Netherlands

Scholarly primitives, such as representing, discovering, sampling and referring are typically well supported by data repositories. However, these do not provide evidence that the data or features provided by the repository are managed and operated in a trustworthy and reliable manner. Assessment against an agreed standard is a way to evaluate the reliability and sustainability of data repositories. Certification helps to demonstrate this to the repository users and funders, thus ensuring them that available data and services are of a certain quality that is defined in corresponding requirements.

The CoreTrustSeal offers baseline certification for repositories and is playing an increasingly prominent role in the ecosystem of generating and serving FAIR data. For CoreTrustSeal certification, a repository first conducts an internal self-assessment against the 16 CoreTrustSeal requirements, which is then reviewed by community peers. Even outside of the formal certification framework the CoreTrustSeal criteria provide a demonstrable approach to internal and external review, supporting a benchmark for comparison and a means to determine the strengths and weaknesses of data repositories.

The main target audience of this workshop are the creators or managers of data repositories who are (planning to) seek the CoreTrustSeal certification or who want to understand how their data is managed in the context of CoreTrustSeal requirements. However, the workshop will be useful to anyone wishing to build digital research infrastructure in a sustainable way or understand the characteristics of such an infrastructure.

Participants of the workshop will have the opportunity to familiarize themselves with the CoreTrustSeal certification requirements and procedures, gaining a better understanding of the necessary steps for certification, and how compliance with the criteria may be met. CoreTrustSeal experts will share their experiences with certification and answer certification-related questions. Participants are asked to take notice of the "CoreTrustSeal Extended Guidance" and

*Speaker

†Corresponding author:tparkola@man.poznan.pl

‡Corresponding author:rene.van.horik@dans.knaw.nl

drafting a self-assessment response to at least 2-3 requirements.

We will also describe why and how the current SSH research infrastructures (CESSDA, CLARIN) use and recommend CoreTrustSeal as means for supporting sustainable FAIR science.

Overall agenda of the workshop:

1. Introduction to CoreTrustSeal
2. CoreTrustSeal approach in CESSDA and CLARIN
3. Case studies / hands-on session

The workshop is part of the certification support activities of the Social Sciences & Humanities Open Cloud (SSHOC) project (<https://sshopencloud.eu/>) that will create the SSH part of the European Open Science Cloud. SSHOC has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782.

Names of organizers:

René van Horik, Data Archiving and Networked Services

Tomasz Parkola, Poznań Supercomputing and Networking Center

List of speakers/instructors:

Birger Jerlehag, Swedish National Data Service

Daan Broeder, KNAW Humanities Cluster/CLARIN ERIC

Rene van Horik, Data Archiving and Networked Services

Tomasz Parkola, Poznań Supercomputing and Networking Center

Keywords: trustworthy repositories, data repositories, CoreTrustSeal requirements

Addressing legal and ethical dimensions of the processing of personal data in research contexts: The ELDAH Consent Wizard as a tool for fostering best practices for consent and transparency in DH.

Walter Scholger ^{*†} ¹, Vanessa Hanneschläger ^{*2},
Koraljka Kuzman Šlogar ^{*3}

¹ University of Graz – Austria

² Österreichische Akademie der Wissenschaften – Austria

³ Institute of Ethnology and Folklore Research (IEF) – Šubićeva 42, 10000 Zagreb, Croatia

One of the core concerns of research in the 21st century is to open up knowledge and science itself. This idea, at the heart of the Open Science movement, encompasses not only the idea of enabling democratic access to research results, but also a careful, responsible and transparent handling of the research process and the research objects themselves, especially when these objects are actually subjects (i.e. people).

The ELDAH Consent Wizard project, co-funded through the DARIAH-EU Working Group funding scheme and the Austrian national CLARIAH-AT funding scheme, set out to develop a "consent form wizard" which enables digital scholars and the wider research infrastructure community to quickly and easily obtain a standardized, GDPR-compliant consent form that is legally valid in all of the European Union, and also sets a best practice for transparency and ethical conduct towards research (and hence data) subjects. Depending on the context for which consent is obtained (e.g. use of images of people, processing of information shared in surveys, collection/processing of personal data), wizard users receive a consent form tailored to their specific needs after answering a series of questions.

In this workshop, participants will be instructed about the basics for approaching such problems. As a first step, an overview of the current legal situation in the area of data protection will be given by presenting the EU-wide General Data Protection Regulation (GDPR) and summarizing the necessities arising from it.

In a second step, the workshop organizers will present the usage of the ELDAH Consent Wizard tool by introducing its functionalities, presenting the predefined usage scenarios and the actual consent forms generated through the tool.

*Speaker

†Corresponding author: walter.scholger@uni-graz.at

Consequently, participants will be able to use the tool themselves to generate consent forms based on their own needs and interests, and test the usability of the resulting consent forms by switching between the roles of data controller and data subject. The participants' feedback will be collected to further improve the development of the ELDAH consent wizard and identify additional community needs.

General structure (incl. rough time estimates):

(15 min) Welcome and introduction of the Consent Wizard Project

(15 min) Introduction round (participants' name, affiliation, consent-related focus)

(45 min) Introduction to European Data Protection laws and the GDPR

(15 min) Q&A on general questions related to legal context

(Break)

(30 min) Introduction of the tool and the usage scenarios

(30-45 min) Interactive testing of the Consent Wizard tool

(30-45 min) Feedback by participants on usability (and usefulness), collection of suggestions for improvement and further development

(15 min) Wrap-up

Target audience: international research community (as users of the tool who wish to tailor consent forms to their research purposes), general public (as potential consentees who will complete consent forms generated with the tool)

Scope: half-day workshop

Instructors: Vanessa Hanneschläger, Pawel Kamocki, Koraljka Kuzman Šlogar, Walter Scholger (ELDAH Working Group)

Keywords: GDPR, Open Science, data protection, tool

Using Collaborative Annotation to Explore and Analyse Uncertainties in Digital Humanities Datasets

Michelle Doran ^{* 1}, Jennifer Edmond ¹, Alejandro Benito-Santos ^{* 2}, Roberto Theron ^{* 2}

¹ Trinity College Dublin – Ireland

² Universidad de Salamanca – Spain

The immediate intention underlying John Unsworth’s explication of scholarly primitives was to suggest some recursive functions of humanities scholarship that might be usefully embodied in tool-building in humanities computing.

The central challenge of uncertainty representation has increasingly come to the fore in Digital Humanities research. The new practices of knowledge production characterised and analysed by Digital Humanities imply uncertainty by their very nature, they are intended to alter our understanding and they almost always occur at the boundaries between different disciplines. From the creation to the consumption of digital resources, there are new stakeholders, contexts and tasks to consider. The amount of digital resources produced (or digitized), stored, explored, and analysed in any Digital Humanities project is immense and uncertainty is a key issue that could hinder substantially in the use and re-use of digital cultural heritage resources. Therefore, traditional humanities methods for managing uncertainty have to be either substituted or aided with ancillary tools and novel user interfaces in this environment. As this exposure to uncertainty can be inherently complicated, especially for users with low computational literacy, the use of appropriate visualization and HCI techniques is key to enhance the comprehensibility of the presented results and to enable a well-informed decision-making process.

This workshop presents the ongoing work of the interdisciplinary PROgressive VIual DEcision-Making in Digital Humanities project (PROVIDEDH), a three-year project funded within the CHIST-ERA call 2016 for the topic “Visual Analytics for Decision Making under Uncertainty – VADMU.” The PROVIDEDH project intends to bring uncertainty to the surface by giving Digital Humanities researchers a space to explore and assess the completeness and evolution of digital research objects, the degree of uncertainty that the models applied to the data incorporate and to share their perspectives and insights with the project’s broad range of stakeholders. The project’s goal is to produce example implementations of uncertainty-aware data analysis pipelines to support Digital Humanities research, putting special emphasis on advancing the state-of-the-art of HCI and visualization in this context. This goal is realisable through the

*Speaker

kind of support and enhancement of the "Functional Primitives of Humanities Scholarship" envisioned by Unsworth, chief among which is the application of collaborative annotation of TEI texts to the explorative analysis of datasets.

In this workshop, we showcase the benefits of our platform by conducting a collaborative experiment that consists of three well-differentiated parts: Firstly, we will ask participants to read and annotate example texts with uncertainty using the proposed interface. After these annotations have been stored in the system, they will be explored using a set of dedicated visualizations to gain insight on the nature of the dataset, e.g. identifying incomplete excerpts, ambiguous references and potential transcription errors, among other traits. Finally, the annotations will be used by the system to support collaborative visual decision-making in a standard entity normalization task typical of many Digital Humanities projects.

Target Audience: Those involved in a Digital Humanities project with special interest in uncertainties, data complexity, and visualization.

Maximum: 20

Keywords: Annotation, Uncertainty, Visualization, Scholarly Primitives

Does Communication belong to Scholarly Primitives? Innovations in scholarly communication through the lenses of Digital Humanities

Erzsébet Tóth-Czifra ^{* 1}, Claire Clivaz ^{*† 2}, Maciej Maryl ^{*‡ 3}, Sara Schulthess ^{§ 2}, Jadranka Stojanovsk ^{*¶ 4}

¹ Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

² DH+, SIB, Lausanne – Switzerland

³ Institute of Literary Research, Polish Academy of Sciences (IBL PAN) – Nowy Swiat 72 00-330
Warsaw, Poland

⁴ University of Zadar – Croatia

- Rationale

Communicating is a par excellence scholarly primitive, all probably the most universal scholarly activity across disciplines and over time. This, however, does not imply the uniformity of scholarly communication practices. This is especially true for the humanities domain where the need for cultural nuance in many fields, scholarly outputs are usually grounded in regional, national and language-specific communities. But what are the chances and limitations of bibliodiversity and media diversity in everyday scholarly communication practices? How we can communicate research results in ways that truly align with our increasingly digital and diverse research workflows? What are the cases where the evaluation is disconnected to novel forms of research and what are the possibilities for re-harmonization? Considering that the traditional paradigm of article and book publishing still serves as the highest value currency of career-development in Humanities, the stakes are especially high when looking for answers to such questions.

The Innovations in Scholarly Communication Work Package of the Horizon2020 project OPERAS-P is conducting an open consultation to get a better understanding about current trends, gaps and community needs in scholarly writing and research evaluation practices. Working closely with the OPERAS infrastructure dedicated to scholarly communication in the SSH domain, we are in the position to directly translate these needs into supporting the development of the relevant OPERAS activities and services. This workshop, organised in coordination with the DiMPO WG, will be centred around the four scholarly primitives of communicating, linking, commenting, and evaluating. We will look how these basic scholarly activities can be supported

*Speaker

†Corresponding author: claire.clivaz@sib.swiss

‡Corresponding author: maciej.maryl@ibl.waw.pl

§Corresponding author: sara.schulthess@sib.swiss

¶Corresponding author: jadranka.stojanovski@irb.hr

by novel scholarly communication practices.

- Workshop format

Target audience: scholars in various humanities disciplines from the DARIAH-community.

What participants can gain: Shared knowledge of workflows and resources for innovative scholarly communication that are well-grounded in research realities of the Humanities. Participants' needs will be amplified in the European policy debates and will be considered in the development of scholarly communication services provided by OPERAS:

Contribution to DARIAH community: Although scholarly communication is an elementary scholarly activity, this topic has not yet been addressed in DARIAH Annual Event workshops. Recommendations will be incorporated into both DARIAHOpen and DARIAH DiMPO outputs, as well as the forthcoming recommendations to Humanities researchers being prepared by the OPERAS Innovations WG.

Method: World café with questions addressing various aspects of research publications (writing, linking, commenting, evaluating) in the humanities. The workshop will build upon the versatile experience of the DiMPO and the OPERAS Working Groups covering such techniques as user stories, process collage, thermometers of indicators etc. adapted for this workshop.

Tentative outline:

- Brief introduction to the scholarly communication workflows, examples on innovative writing and co-creation practices, the many flavours of openness in peer review etc. ; participatory knowledge creation, design thinking and problem crowdsourcing on the given case study.
- Core questions are identified through the participatory process. In a world café setting, participants divided into tables are discussing and presenting their outcomes.
- In the summary phase the core challenges are identified and addressed in more detail.

Keywords: scholarly communication, writing practices, peer review, Digital Humanities, research evaluation, multimedia, multilingualism

Bibliographical Data Workflows: discover, analyse, and improve them

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

Organizers: DARIAH ERIC "Bibliographical Data" Working Group

Type of workshop: 4-hour workshop

Instructors:

Vojtěch Malínek

Tomasz Umerle

Matteo Romanello

David Lindemann

Abstract:

Bibliodata involves us all

Bibliographical data serves to identify documents crucial for the humanities - from books, journals, or articles, to grey literature and web-resources. Although the scope of bibliodata assigned to documents varies, there are some commonalities, like the attribution of authorship, physical description, and subject classifications. Concurrently, bibliodata is enriched with, e.g., citation data, linked data, persistent identifiers, social tags, annotations.

All researchers perform bibliodata activities: we (co)create, and (re)publish documents, (co)produce bibliographical descriptions, (re)use them, or depend on citation metrics. Also, we all take advantage of the services which depend on bibliodata quality, from library catalogues, through scholarly repositories, to bibliodata managers. Lastly, bibliodata research has heavily influenced DH landscape, with contributions in literature (F. Moretti, M. Jockers, K. Bode), intellectual and book history (S. Burrows, A. Montoya, M. Tolonen), correspondence network studies (H. Hotson, D. Edelstein, T. Wallnig) or citation studies (S. Peroni).

Bibliodata workflows and scholarly primitives

*Speaker

†Corresponding author:tomasz.umerle@ibl.waw.pl

Collaborative bibliodata workflows are repeatable patterns of co-dependent activities involved in bibliodata curation, and research.

Analysis of research activities that constitute those workflows is crucial to discussions about contemporary set of scholarly primitives, and the relevance of the notion itself.

During our workshop we want to investigate bibliodata activities performed by the audience on a day-to-day basis, and highlight their role in the research workflows. We will try to assess whether the actual bibliodata activities are well-supported by existing bibliodata research infrastructure, but also discuss the nature of current bibliodata RI's limitations.

Workshop's goals

We aim to:

- present basic information on bibliodata,
- survey the audience's research bibliodata activities, and investigate their role in workflows,
- pinpoint inefficiencies in bibliodata workflows through critical analysis of existing patterns.

Preliminary programme:

Introduction:

- what is bibliodata, how they are structured and used,
- what are the main bibliodata stakeholders, and research approaches,
- what are the main bibliodata activities in DH?

Audience survey:

- what bibliodata they work with and what activities they engage in,
- which tools, and services they use - which stakeholders and research approaches they interact with,
- how bibliodata activities relate to audience's research, what value and meaning they assign to it.

Teamwork sections with short introductions before each one:

- Investigating workflows that involve researchers as bibliodata producers;
- examples: providing metadata in scholarly repositories,

co-managing and sharing bibliodata collections,

compilation of course bibliographies for students

Investigating workflows that involve researchers as bibliodata re-users;

examples: patterns of bibliodata re-use in digital projects, discovering, and

publishing bibliodata.

Based on critical investigation of identified workflows we would like to think outside the box: what are the workflows that are not developed enough? what we are missing out on?;

examples: new data, data accessibility, missing metrics, data quality, lack of, or mishandling of existing tools?

Target audience

Our workshop will move back and forth between short presentations on bibliodata, based on concrete examples from the humanities, and interactive teamwork, so no previous experience of bibliodata is expected, and no in-depth expertise is needed to participate in teamwork during our workshop.

Keywords: bibliographies, bibliographical data, metadata, data analysis, data quality assessment

Information Extraction Workflow for Digitised Entry-based Documents

Mohamed Khemakhem ^{*† 1,2,3}, Simon Gabay ^{*# 4}, Béatrice Joyeux-Prunel^{§ 5}, Laurent Romary ^{*¶}, Léa Saint-Raymond ⁶, Lucie Rondeau Du Noyer

¹ Inria - ALMAAnCH – Inria-ALMAAnCH – France

² Centre Marc Bloch – Germany

³ Université Paris Diderot - Paris 7 – Université Paris Diderot - Paris 7 – France

⁴ Université de Neuchâtel – Switzerland

⁵ Université de Genève – Switzerland

⁶ Ecole normale supérieure, Paris, DMA – Ecole Normale Supérieure de Paris - ENS Paris – France

1. Program

It is a 4 hours workshop:

- 30 mn: introduction of the workflow and the two tools
- 30 mn: setting up the tools on the machines
- 2 h: Two Parallel groups. Goal: Training a model for a dictionary and a model for a catalogue
- 30 mn: towards a digitisation pipeline
- 30 mn: Q&A

2. Instructors

- Mohamed Khemakhem
- Simon Gabay
- Laurent Romary

3. Target Audience

*Speaker

†Corresponding author: medkhemakhemfsegs@gmail.com

#Corresponding author: simon.gabay@unine.ch

§Corresponding author: Beatrice.Joyeux-Prunel@unige.ch

¶Corresponding author: laurent.romary@inria.fr

Scholars and humanists with basic notions of command lines and familiar with the TEI standard.

4. Abstract

The massive retro-digitisation of legacy paper resources in the last decade, along with the constant growth of compiled unstructured digital text material, have created an unbalanced situation where the existent ad hoc techniques for exploiting such resources are unable to cover the important stream of emerging corpora. In this workshop we address this issue and present an exploratory workflow implemented in two state of the art infrastructures for Information Extraction (IE) from documents with entry-based structure and diverse content.

IE in Digital Humanities (DH) has always been a serious challenge for researchers dealing with modern or legacy text resources [1, 2]. GROBID-Dictionaries is a project which has been launched to fill in this gap by accelerating the modelling and structuring of resources within the lexicography field. The first version of the machine learning infrastructure has been focused on structuring digitised dictionaries into TEI-compliant resources [3].

In GROBID-Dictionaries, the activation of cascading IE models follows an exploratory process based on the MATTER workflow [4]. Throughout a multi-stage annotation and curation process, a user of the tool discovers gradually the structure and the variation of the information in a target document.

In addition, and despite being initially designed for structuring dictionaries, the implemented approach and the tool's pluggable models have shown enough flexibility to be applicable on a wider range of modern and legacy entry-based documents such as dictionaries [3, 5] (fig. 4), legacy address directories [6] (fig. 6), and large bibliographic collections [7].

The rising interest around the applicability of such an IE approach to certain entry-based documents [8, 9] and its fitness potential in larger document processing pipelines, triggered the creation of a second tool adapted for analysing and structuring a family of prints: Catalogues (fig. 5 & 7).

Grobid-Cat is a fork of Grobid-Dictionaries that remains faithful to the exploratory cascading approach with more optimised combination of parsing models (fig. 2) and more adapted TEI encoding (fig. 3).

The proposed workshop represents an illustration of scholarly primitives practices employing a new generation of IE tools. Our goal is to familiarise interdisciplinary DH users with the MATTER workflow through two parallel hands-on sessions, each one dealing with a different category of digitised documents: dictionaries and catalogues. We rely on enhanced usability features [5]

to ease the setup and the manipulation of the two systems for relaxed required IT skills. Besides presenting new powerful means for manipulating, exploring and extracting structured text from digitised material, the workshop gives DH researchers the opportunity to have an idea about the challenges of integrating such tools in digitisation pipelines. This includes providing background on the different factors interfering in the performance of machine learning models, such as information modelling, text encoding, OCR system choice and annotation consistency.

Keywords: Information Extraction, Digitised Documents, Entry based documents, dictionaries,

catalogues, machine learning

Poster/demo session

CLARIAH-DE - Aligning two research infrastructures: Experiences and Challenges

Stefan Buddenbohm ^{*†} ¹, Dirk Goldhahn ²

¹ Göttingen State and University Library – Germany

² Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig – Germany

CLARIAH-DE is the merger of the two research infrastructure networks CLARIN-D and DARIAH-DE. CLARIAH-DE is merging the two established research infrastructures over the course of two years (2019-2021). Thus, academic endeavours of researchers from the humanities and cultural sciences who work with complex digital tools and special data repositories shall be substantially facilitated. Both research infrastructures pursue goals directly connected to the scholarly primitives - discovering, annotating, comparing, referring, sampling, illustrating and representing - as they provide services and resources for them.

A large work package (WP4) is devoted to the harmonisation of the CLARIN-D and DARIAH-DE technical infrastructures. Infrastructure has to be seen both as basic or generic components like AAI, PIDs, monitoring, collaboration platforms but also as research oriented services, like the CLARIN Federated Content Search, the DARIAH Collection Registry or a GeoBrowser. Our work is based on previous cooperation, for instance the joint Technical Advisory Board (TAB), and preliminary integrative endeavours in recent years.

But despite well established communication processes such as the TAB it doesn't come as surprise that the practical merger leads to considerable challenges. CLARIN-D and DARIAH-DE have different disciplinary traditions, and have thus developed different technologies, tools, services and processes that can now complement each other. The consolidation, however, is not straightforward, as is the case, in particular, with the merging of the three search and retrieval tools: Generic Search, Federated Content Search and Virtual Language Observatory. Consolidation often also requires the harmonisation of standards and interfaces (resource metadata, interchange formats). In other areas, such as Authentication and Authorization Infrastructure (AAI), we already found a solution based on CLARIN's and DARIAH's existing AAI-approaches. The work package is complemented by a technology watch that goes beyond CLARIN and DARIAH and tries to incorporate the most important developments in the field into its own plans wherever possible.

We must also consider the European level. Both CLARIN-D and DARIAH-DE are contributing partners to the ERICs and limitations that have to be considered in the merger of two national infrastructures come with this interwovenness. CLARIAH-DE is directly involved in European development through its two partners. CLARIN-D is part of CLARIN - European Research Infrastructure for Language Resources and Technology and DARIAH-DE is the German partner of Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU).

*Speaker

†Corresponding author: stefan.buddenbohm@sub.uni-goettingen.de

In addition, and as research driven infrastructures, the community perspective is possibly most important, joining the picture of interdependencies: both CLARIN-D and DARIAH-DE are well established with users and dissemination activities which contributed to the level of trust researchers have in them. Trust is a very important and delicate asset which has to be taken into consideration at all stages of the merger process.

The poster presents some of our experiences in the merger process. We try to reflect not only on an infrastructural level but also involve the community perspective. We want to use this opportunity to collect feedback particularly from the European partners and projects and are interested in useful experiences from other initiatives.

Keywords: CLARIAH, DE, CLARIN, D, DARIAH, DE, merger, infrastructures, Germany, poster

CLARIAH-DE - Added value of the infrastructure merger

Stefan Buddenbohm * ¹, Thorsten Trippel ², Anne Klammt ¹, Dominique Rose ², Roberta Toscano ²

¹ Göttingen State and University Library – Germany

² University of Tübingen – Germany

CLARIAH-DE brings together the data, tools and services of the two research infrastructure initiatives CLARIN-D and DARIAH-DE and the existing communities of users and providers of the infrastructures until 2021.

CLARIN-D and DARIAH-DE can partly be seen as German equivalent of the European research infrastructure initiatives CLARIN and DARIAH. Potential users being aware of CLARIN and DARIAH, particularly users outside of the in-group, often wonder about the differences between both infrastructures and services. Most potential users want to access services and resources in the most convenient way, preferably with a single point of access, and want to use the services and resources they need, independent of their origin and also in combination. CLARIAH-DE also intends to open up resources for other communities and to extend its user base.

During the merger process, CLARIAH-DE faces challenges at several levels:

- (Technical level) Different approaches towards architecture, standards and services have to be considered. A merge or complementary use is not a realistic goal with look at all resources. On a technical level, different systems and rules of participation were part of the initiatives, for example with certification processes of data and service providers.
- (Communication level) Both infrastructures have developed different cultures of providing resources and of communicating with their users. For instance, dissemination and training have to be synchronized.
- (User level) The established userbase may benefit of the merger but need to become aware of the options which aligns with the need to extend the user base.

One of the most delicate assets that has to be considered is merging the processes. This includes the process of accepting data, tools and services as part of the merged infrastructure, but also the processes in the governance structures. It also touches on the trust that the users have in the research infrastructures and that should ideally be transferred to CLARIAH-DE. For this end, we will address various aspects through our work in the project:

*Speaker

- Besides a lean project governance, five work packages have been designed, each lead by a tandem of partners from both originating infrastructures. Each work item was designed with the requirement that both infrastructures would be contributing and that the result would provide added value to the existing users of each underlying service.
- The WPs also provide a platform for discussions on data formats, technical infrastructure, integration of major tools, training and teaching integration, and dissemination.

After a year of collaborative work in CLARIAH-DE and also due to proceeding coordination efforts between CLARIN-D and DARIAH-DE, we have achieved first results: we were able to develop a common language and an understanding of the respective other infrastructure. The partners from different contexts have been successful in establishing a productive working groups. The integration of basic infrastructures is headed in the right direction. A first joint website is available and in the future becoming a single point of access to the data, tools, and services provided originally by the two infrastructures.

Keywords: CLARIAH, DE, CLARIN, D, DARIAH, DE, merger, infrastructures, Germany, poster

The SSH Open Marketplace: How to contribute to this discovery portal

Stefan Buddenbohm *¹, Laure Barbot *², Frank Fischer *³,
Tomasz Parkola *⁴

¹ Göttingen State and University Library – Germany

² Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

³ Digital Research Infrastructure for the Arts and Humanities – Germany

⁴ Poznan Supercomputing and Networking Center – Poland

The SSHOC cluster project (Social Sciences and Humanities Open Cloud) is a contribution to the European Open Science Cloud (EOSC) and addresses particularly the requirements of the Social Science and Humanities research community. An important goal of the project is to deliver the SSH Open Marketplace (<https://sshopencloud.eu/marketplace>) as a discovery portal and a comprehensive information hub for the SSH research community regarding tools, services, training materials and other resources. Beyond this straightforward approach of identifying resources the marketplace also utilises the research process as an underlying matrix, e.g. in addressing research methods as a composition of actions linked with resources, and where possible, connect them with tutorials, feedback or assessment functions.

This sketch of the SSH Open Marketplace obviously connects very well with the scholarly primitives as described by John Unsworth – discovering, annotating, comparing, referring, sampling, illustrating, and representing – as they shape the requirements the Marketplace has to meet to be of actual value for the research community. This also aligns with the DARIAH Strategic Plan 2021–26 which stands upon four pillars: (1) building the Marketplace as a humanist-friendly component of the European Open Science Cloud, (2) approaching training and education strategically and in a coordinated fashion, (3) deepening our connection to our communities and ensuring they are with us, and (4) strengthening our voice in policy and advocacy.

A testing instance of the marketplace is available here: <https://sshoc-marketplace.acdh-dev.oeaw.ac.at/> SSHOC follows an agile development approach with user feedback from the very beginning and feedback loops in between. The audience of the DARIAH Annual Event – humanities scholars and the DARIAH community – is a perfect multiplier for us to receive feedback from the research community. Beyond this we invite the DARIAH community at large, members of the DARIAH working groups, NCCs, students and researchers. In short: Does the SSH Open Marketplace

*Speaker

meet the requirements of the DARIAH communities from the Arts and Humanities SSH community in terms of content and functions and if not entirely, what is missing? This question particularly addresses the sufficiency of the Marketplace in terms of its content and usability. The poster will present the key components of the SSH Open Marketplace and the context of its development in the SSHOC project, while the demo will be an opportunity to have an interface walkthrough at the portal, its functions and its contents and to encourage participants to provide their feedback.

Keywords: SSHOC, Marketplace, DARIAH, SSH Open Marketplace, agile development, feedback, user requirements, community

The DARIAH ELDAH consent form wizard

Vanessa Hanneschläger ^{*} ¹, Walter Scholger ^{*†} ², Koraljka Kuzman
Šlogar ^{*#} ³

¹ Österreichische Akademie der Wissenschaften – Austria

² University of Graz (Uni Graz) – 8010, Universitätspl. 3, 8010 Graz, Austria

³ Institute of Ethnology and Folklore Research (IEF) – Šubićeva 42, 10000 Zagreb, Croatia

In this poster and demo, we would like to introduce the "consent form wizard" developed by the DARIAH-EU working group "Ethics and Legality in Digital Arts and Humanities" (ELDAH).

This tool is currently under development and will be finalized and available in time for the DARIAH annual meeting. It will enable digital scholars as well as the wider research infrastructure community to quickly and easily obtain a standardized consent form that is conforming to the obligations and regulations of the European Union General Data Protection regulation (GDPR) and therefore legally valid in all of the European Union.

Depending on the context for which consent is obtained (e.g. use of images of people at academic events, processing of information shared in surveys, collection / processing of personal data), the wizard users will receive a consent form tailored to their specific needs after answering a series of questions. The different scenarios are the result of discussions with participants of the DH2019 in Utrecht and the DARIAH Annual Meeting 2019 in Warsaw and hence driven by the needs of the research community. Additional meetings with stakeholders from cultural heritage organisations (e.g. ICARUS) and the DARIAH-EU community will ensure a critical evaluation as well as a pragmatic, user-friendly implementation of the tool.

It will build on the code of the CLARIN LINDAT Public License Selector which offers the same service for license selection that we would like to develop for obtaining a consent form. Software developers, legal experts and (digital) humanities researchers will cooperate in the development of this tool, which - while legally addressing the strict data protection regulations of the European Union - will also be of use for the international research community outside Europe, since the consent forms will provide a best-practice template for ethical research conduct when processing personal data, and hence address the increasingly prominent topic of ethical research practice and scientific behaviour, especially in a largely digital, internet-based research context.

The presentation of this poster and demo at the DARIAH annual event will allow us to present the wizard to our infrastructures colleagues, discuss their needs and experiences with the protection of personal data and hence, if appropriate, adapt the wizard to reflect European common perspective on research ethics and the processing of personal data.

DARIAH ELDAH blog. <https://eldah.hypotheses.org/>

*Speaker

†Corresponding author: walter.scholger@uni-graz.at

#Corresponding author: koraljka@ief.hr

General Data Protection regulation (GDPR). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:>

International Centre for Archival Research. <https://icar-us.eu/>

Kamocki, Pawel, Pavel Stranák, and Michal Sedlák. "The Public License Selector: Making Open Licensing Easier." Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, et al., 2533–2538. Paris: European Language Resources Association (ELRA), 2016. http://www.lrec-conf.org/proceedings/lrec2016/pdf/880_Paper.pdf

Markham, A., Buchanan, E. (2012): Ethical Decision-Making and Internet Research Recommendations from the AoIR Ethics Working Committee. <http://aoir.org/reports/ethics2.pdf>

Keywords: Consent Forms, Legal Consent, Legal Issues, Data Privacy, consent form wizard

Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism - The State of the Art

Sanja Peric Gavrancic ^{*† 1}, Martina Kramaric ^{*# 2}

¹ Institute of Croatian Language and Linguistics – Croatia

² Institute of Croatian Language and Linguistics (IHJJ) – Republike Austrije 16 10 000 Zagreb, Croatia

Retro-digitization of Croatian linguistic heritage is an important component of the Digital Humanities Network in Croatia, as an area of scholarly activity at the intersection of the computing or digital technologies and the disciplines of the humanities. Despite that, the creation of retro-digitized resources in Croatia is still in its beginnings and the existing resources do not include grammars from the pre-standard period of the Croatian language. Therefore, there is no repository that would contain a digitized and searchable database of Croatian language grammar books of the period, no model that could be applied and no systematically conducted scholarly research in the field. The aim of the project *Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism* to change this situation.

This innovative (in the context of the European and Croatian philology) and comprehensive approach of presenting eight historical grammars implies the transfer of the printed media to computer-readable and searchable text. In this project, it also includes a multilevel mark-up of the transcribed or translated grammar text and its connection with facsimiles. Digitization process will include the mark-up of the grammar text segments and morphological paradigms by TEI tags and will be conducted on transcriptions or translations of the selected grammars. The way of grammar editing depends on the particularities of each grammar, as not all grammars are written in Croatian, nor all grammars describe the Croatian language.

The main aim of the project is to create a web portal of the Croatian grammar books before Illyrism (before the choice of the common standard language and orthography), which would include facsimiles of selected grammar books with basic information, transcription or translation and an index of historical grammar and linguistic terminology. The portal will be equipped with thematic searching possibilities on the morphology level.

This innovative and digital approach of presenting old texts (which are not always easily accessible) fundamentally changes the perspective and practicalities of conducting research on the historical material. Digital facsimiles will enable research on the material which was not presented before (some grammars are still in manuscripts) and the individual mark-up of the morphological paradigms will enable comparison between the selected grammars.

The web portal of early Croatian grammar handbooks will be open for the general public and

*Speaker

†Corresponding author:speric@ihjj.hr

#Corresponding author:mkramar@ihjj.hr

will serve as an example of presenting historical texts of any kind and as a starting point of similar research in the different fields of humanities, and particularly historical linguistics. This kind of digital infrastructure is available for further upgrade and opens the possibility of linking and exchanging knowledge with other institutions.

This work has been fully supported by Croatian Science Foundation under the project *Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism* IP-2018-01-3585.

Research group: dr. Petra Bago, dr. Martina Kramarić, dr. Ivana Lovrić Jović, dr. Ana Mihaljević, dr. Sanja Perić Gavrančić, dr. Ivo Pranjković, dr. Diana Stolac, dr. Ljiljana Šarić, dr. Barbara Štebih Golub, dr. Tamara Tvrtković. The project leader: dr. Marijana Horvat.

Keywords: retro, digitization, prestandard Croatian grammars, linguistic heritage, XML, TEI, digital humanities

TaDiRAH as Linked Open Data

Canan Hastik * ¹, Luise Borek * ¹

¹ Technical University of Darmstadt – Germany

Classifying and categorizing the activities that comprise the Digital Humanities (DH) has been a longstanding area of interest for many practitioners in this field, fueled by ongoing attempts to define Digital Humanities both within the academic and public sphere. The emergence of directories that cross traditional disciplinary boundaries has also spurred interest in categorization, with the practical goal of helping scholars identify, for instance, projects that take a similar technical approach, even if their subject matter and objects are vastly different.

TaDiRAH, the Taxonomy of Digital Research Activities in the Humanities, is the result of a year-long project undertaken by the DiRT (Digital Research Tools) Directory and DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities) to develop a shared taxonomy. TaDiRAH was created to organize the content of sites as diverse as the DARIAH Zotero bibliography ‘Doing Digital Humanities’, the DHCommons online hub and DiRT (now TAPoR, the Text Analysis Portal for Research). It has also been used with the Standardization Survival Kit research use case scenario collection, and AGATE, a European Science Academies Gateway for the Humanities and Social Sciences, to name just a few examples.

TaDiRAH has been shaped both by and for the DH community to classify and define research activities in the humanities and related fields. With DHCommons, DARIAH and Europeana as its original use case environments, it has developed into a widely known taxonomy that has been used and adapted in various international research projects. Through a community-driven approach, TaDiRAH has been improved through user feedback and has been translated into French, German, Spanish and Serbian increasing its application. However, to meet an increasing demand that is intensified by the growing influence of linked data technologies ensuring interoperability in digital research infrastructures, TaDiRAH should now evolve from a sum of terms in the form of narrower and broader relationships to a formalized model with a common ontological basis. Within CLARIAH-DE the taxonomy is now translated into a machine-readable version in order to implement it in the Language Resource Switchboard (LRS). This process includes the conceptualization, semantification and formalization of the existing taxonomy in terms of the FAIR Principles. It will be made available as Simple Knowledge Organization System (SKOS) including a SPARQL endpoint becoming part of DARIAH-EU’s Vocabs services hosted at the Austrian Centre for Digital Humanities Austrian Academy of Sciences (ACDH-ÖAW).

TaDiRAH’s aim remains to be a community-driven taxonomy that is easy to use and meets the needs of a wide variety of humanities scholars. The goal, therefore, is to provide a version that is more fully compliant with standards for Linked Open Data (LOD). With several European initiatives currently shaping advanced research infrastructures that could benefit from

*Speaker

its implementation, new collaborations have been formed to bring TaDiRAH to this next level. TaDiRAH's implementation may lead to the taxonomy living on and help categorize, visualize, search, and find the activities and results of Digital Humanities initiatives.

Keywords: Taxonomy, Digital Humanities, Community, LOD

Digital scholarly primitives : how data curation/implementation/aggregation and mapping happen in ArkeoGIS

Loup Bernard * 1

¹ Université de Strasbourg (UMR 7044) – université de Strasbourg – MISHA 5 allée du Gal Rouvillois CS 50008 67083 Strasbourg cedex, France

ArkeoGIS is a multidisciplinary application. The databases come from different sources: institutional researchers' work, graduate students' researches, private companies and archaeological heritage management services. They also come from palaeoenvironmentalists', geographers' and historians' work. Multidisciplinarity is spurred and promoted All these shared databases are available and can be queried through the webGIS by ArkeoGIS' users.

Each user have its own personal and customizable project interface. One can query online all or only a part of these databases, display the results on multiple backgrounds maps, save and export them towards other tools one might use (CSV export).

Actually in release 4.2 the poster will present the latest evolutions of the tool (multichronology and shape display), and encourage users to create new communities using the platform but also discuss how the scholarly primitives evolve with digital tools.

What is ArkeoGIS ?

ArkeoGIS enables to pool and query – thanks to a mapping interface – the spatialized scientific data about the past (archaeology, environment...). ArkeoGIS is an online platform accessible in four languages (German, English, Spanish and French).

ArkeoGIS is a multidisciplinary application. The databases come from different sources: institutional researchers' work (either personal or contractual researches), graduate students' researches, private companies and archaeological heritage management services. They also come from palaeoenvironmentalists', geographers' and historians' work. Multidisciplinarity is spurred and promoted All these shared databases are available and can be queried online by ArkeoGIS' users.

Each user have its own personal and customizable project interface. One can query online all or only a part of these databases, display the results on multiple backgrounds maps, save and export them towards other tools one might use (CSV export).

The chronological frame of the tool is now both customizable and multiple. ArkeoGIS' multi-chronologies system enables to aggregate, for different areas, information from the Prehistory to the Present. About ten chronologies are already available, from the Iberian Peninsula and the Mediterranean to the Middle East or continental Europe. ArkeoGIS' geographical frame enables to display information about all of these regions. As today, the most documented areas are the Upper Rhine Valley, the Mediterranean and the Middle East.

*Speaker

ArkeoGIS can be used for many different forms of research, individual or collective. It allows, among other things, to handle the data management plan (DMP) for contractual researches. ArkeoGIS is a powerful tool to use during different kind of studies (excavations, synthesis or PhD thesis, etc.).

Several tens of thousands of sites, objects and analysis are already available. ArkeoGIS is also linking several digital tools, allowing its users to be aware of their existence.

Every author submitting his localized data into ArkeoGIS keeps the control over them and is the only one who can amend them. Any user can easily access to other contributors' data and improve its own database. A directory allows researchers to contact each other. This initiative helps to develop scientific exchanges between countries and institutions.

Keywords: archaeology / online mapping / platform / rawdatashare / LOD / FAIR

Education as a Scholarly Primitive - The DH Course Registry

Hendrik Schmeer ^{*† 1}, Tanja Wissik ^{*‡ 2}

¹ CLARIN ERIC – Netherlands

² Austrian Center for Digital Humanities (ÖAW) – Austria

The Digital Humanities Course Registry's (DHCR) main aim is to foster teaching of digital practices by providing a searchable platform for available courses, workshops and education programmes among the Humanities. Data recorded for more than five years about DH teaching provides a sample of the relevance of those scholarly primitives in teaching activities.

The DHCR records are tagged using TaDiRAH [1] categories. While the "Techniques" category, closely related to digital primitives or practices, can still be found in the DHCR metadata, the broader "Activities" category has been dropped in the early days of the registry, as those scholarly "Activities" rendered being rather unspecific to DH teaching.

Metadata about teaching activities in the DH Course Registry allows for metrics along the techniques category alone, or combined for specific objects, disciplines, in specific regions or their development over time. Retrieval of the metadata is now possible for everybody by the new DHCR-API, which has been implemented in spring already, but officially released, announced and documented in autumn 2019, aside the rewritten public web-UI.

This poster demo wants to encourage the use of the public API for more specific analysis, and provides some example analysis. Also, it asks for continuing contributions to the data pool the DHCR is collecting since 2014.

Questions towards a new classification of scholar activities in DH also are of great importance to DH teaching, as well as in the DHCR this reflects, what "DH" is or might be regarded to be. [1] TaDiRAH - Taxonomy of Digital Research Activities in the Humanities,

Keywords: DHCR, Course Registry, DH Course Registry

*Speaker

†Corresponding author:mail@hendrikschmeer.de

‡Corresponding author:tanja.wissik@oeaw.ac.at

The database of antique archaeological sites of the Republic of Croatia

Kristina Turkalj* ¹, Ivana Ožanić Roguljić ^{†‡}, Kristina Jelinčić Vučković

¹ Institute of Archaeology – Croatia

The project *”The database of antique archaeological sites of the Republic of Croatia”* is outgoing project of the Institute of Archeology and it is set on website of the Institute (<http://baza.iarh.hr/public/local>). We intend to present an outgoing project; explain what is *”The database of antique archaeological sites”* and what information can be find in this base.

The Database (BAZA) uses as its foundation the wealth of Croatian archeological sites. It was created in order to structure the data collected during the work of the Institute. As a source of data, the available literature is used together with knowledge gathered from the Institute’s scientific activities and data stored in the archives of the Institute of Archeology. The aim of the project is creating a database of Greek and Roman archaeological sites as the foundation for science, culture, preservation of Croatian heritage, as well as the presentation of cultural tourism. In addition, goals is to educate people about the historical importance of the area, its heritage and to draw attention at the long-abandoned and forgotten sites.

The DATABASE is based on archeological sites in Croatia that are known in the literature or established through field research with the aim of creating a tool that provide easier access to data needed for future scientific research, heritage management, and the creation of professional and scientific projects that have archeological basis heritage.

The database is enriched with information on archeological sites that are attractive enough to become a tourist destination. Data availability enables business, scientific and private entities from Croatia as well as around the world to access information about the desired cultural archaeological destination.

Keywords: Database, antique, archaeology, sites, Croatia

*Corresponding author:kristina.turkalj@iarh.hr

†Speaker

‡Corresponding author:iozanic@iarh.hr

Supporting Decision-Making in the Entity Normalization Task. On the Example of PROVIDEDH.

Grzegorz Bryk ^{*† 1}, Alejandro Benito ², Michał Kozak ^{# 1}, Cezary Mazurek ¹, Szymon Pl-aczaszek ¹, Roberto Theron ²

¹ Poznań Supercomputing and Networking Center – Poland

² University of Salamanca – Spain

The main focus of the PROVIDEDH project is geared toward the creation of visualization systems that aid in the handling of uncertainty in Digital Humanities (DH) datasets. Our tools on the collaborative platform were developed against the background of human-centered design with a focus on easing uncertainty annotation and visualization, promoting the use of TEI standards and making uncertainty play a more active role in the research process.

At the moment our platform allows users to create research projects, upload batches of TEI files, annotate these files concurrently by multiple users as well as identify and manage the entities appearing in the files. Users are able to create their own taxonomies of uncertainty or use a default one to make annotations. All annotations and unifications made on the platform can be rendered then to the TEI format. The user does not need to be familiar with TEI, as our platform provides him/her a friendly interface for annotating electronic texts. This allows a variety of users to use a platform in ways suitable to their various needs and datasets.

We are currently working on visualising existing links between entities in different project files, and providing visual tools for analysing uncertainty, its sources, meaning, and context. Those visualisations should allow users to interact with data in a more intuitive way and see more clearly the connections, trends and anomalies.

In addition to functionalities explained above, enabling users to manually make decisions about the dataset, we plan to incorporate into the platform the Decision Support Subsystem (DSS), which should suggest users actions to be taken with regard to project data. Despite it is planned to support many different exploration scenarios in the future, the system is first showcased to support a typical DH task: entity normalization. In this regard, the system employs a neural network and a weighted multigraphs to automatically recognize entities appearing in the TEI files under different identifiers and to suggest unifications when certain similarities in the input data (connections between nodes in the graph) are detected. In this graph, the nodes (representing entities in the system) are connected by edges constructed and weighted by metadata analysis. For example, the data to be analysed may consist of, inter alia, gender, dates, mentions' context, geodata derived from place names or mutual acquaintances. The intermediate edge-weighting step allows adjusting the impact of different types of connections between the entities in an interactive manner. Additionally, the algorithm learns from users' actions on the

*Speaker

†Corresponding author: grzegorz.bryk@technowacje.pl

#Corresponding author: mkozak@man.poznan.pl

same and other datasets. All this information is progressively presented to the user to inform the decision-making in the normalization task. The aim of this presentation is to familiarize attendees with the current capabilities of the platform, and the development roadmap of the platform until the end of the project. During the conference, we expect to gather important feedback from the DH community in order to advance the state-of-the-art in novel and useful ways in the future.

Keywords: Visual Analytics, Entity Normalizations, Decision Support, TEI, Annotations

DARIAH contributions - visual analytics and workflows

Andrea Scharnhorst *^{1,2}, Femmy Admiraal³, Dirk Roorda¹

¹ Data Archiving and Networked Services – Netherlands

² Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

³ Data Archiving and Networked Services (DANS) – Anna van Saksenlaan 51 2593 HW Den Haag, Netherlands

This paper builds on earlier analysis of the DARIAH in-kind contributions as presented 2019 at the DARIAH Annual Event [1]. Since then, the tool has undergone a refurbishment - moving it from the development phase (after HaS) to a stable service. In parallel, workflows have been streamlined, training sessions and documentation has been established, and the NC's have embraced these changes with new submissions in an engaged and regular way. In the present paper, we report about the current state of the process and present a visual analytics of an extended set of contributions. We attempt a linking to contributions to the different stages of research practices as described in the Scholarly Primitives. We link the empirical analysis of the contributions to the higher-level discussion on how to order, understand, and support the contributions of DARIAH members to the DARIAH European Infrastructure. [2] The contribution is also suitable as a poster.

Scharnhorst, A, Admiraal, F & Roorda, D 2019, 'Visual Analytics of the DARIAH in-kind contributions'. Extended Abstract, DARIAH Annual event "Humanities Data" Warsaw, 2019
Anonymous. (2019). DARIAH Digital Research Infrastructure for the Arts and Humanities - Strategic Plan 2019-2026. Berlin. <https://www.dariah.eu/wp-content/uploads/2019/08/Strategic-Plan-2019-2026.pdf>

Keywords: services, contributions, visual analytics

*Speaker

Communicating: the forgotten, yet indispensable scholarly primitive and its infrastructural dimension

Maciej Maryl ^{*† 1}, Marta Błaszczńska ^{*1},

Agnieszka Szulińska ¹, Paweł Rams ¹, Piotr Wciślik ¹

¹ Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences – Poland

John Unsworth (2000) proposed a tentative list of scholarly primitives, and although he made a reservation that it was not meant to be exhaustive, one omission is striking, namely the exclusion of *communicating*. It is even more visible once one realises that all the examples he actually provides in the paper - *comparing* (IBabble), *linking* (Blake Archive), or *sampling* (VRML visualisation of Dante's Inferno) - have the indispensable communication component. The aim of this presentation is two-fold. First of all, we will reclaim the role of communication as one of the fundamental functional primitives, crucial in all stages of the research workflow. To use Unsworth's nomenclature, communication takes advantage of the additive characteristics of scholarly primitives and enters into combinations with all other scholarly primitives. Secondly, right after reestablishing the communication as a scholarly primitive we will swiftly proceed to problematise the notion of its universality for all disciplines through the exploration of the specificity of scholarly communication in the humanities. We will achieve that using New Panorama of Polish Literature (NPLP.pl) as a case study and the relevant, discipline-specific digital infrastructure for scholarly communication in the humanities.

It has long been suggested that communication should be seen as a fundamental element of the research workflow, rather than an activity running somewhat separate to the research practice (Latour and Woolgar 1986; Garvey 1979; Galison and Galison 1997; Nielsen 2011). Recently this idea was reinforced by Hillyer et al. (2017) who describe open science as "opening of the entire research cycle" and include communication as one of its key elements. It means that dissemination is no longer perceived as the final stage of a research process but becomes an integral part of all scholarly activities. New digital methods and tools (Dallas et al. 2017), including electronic communication and social media (Kjellberg 2010), facilitate this process, allowing scholars to communicate and collaborate with each other and the wider audience quickly and efficiently at all stages of their work. This also includes intermediary results of the work, including raw and secondary data (Castelli, Manghi, and Thanos 2013).

The incorporation of communication into all stages of the research workflow also means that choosing a certain communication strategy is obviously influenced by the perceived goal, but

*Speaker

†Corresponding author: maciej.maryl@ibl.waw.pl

also this very goal has an impact other phases of the research process (e.g. collecting, sampling or linking practices). This feedback loop will be discussed in greater detail on the example of NPLP, a research infrastructure for literary scholars, which enables the creation of extended, multimedia monographs. The scholarly arguments are presented through linking text with image, visualisation, map and video content. Yet, the very act of creating a new digital collection also forces researchers to rethink how their work is presented, categorised and displayed. For instance in "Postmodern Sienkiewicz" collection (<http://nplp.pl/en/kolekcja/postmodern-sienkiewicz/>) authors divided their articles into shorter fragments with additional iconography allowing for non-linear reading and access through image-interface. These activities required additional work on the stage of data collection, analysis and interpretation.

In conclusion, we will tackle upon the question to what extent such communication practices are universal for all sciences and what could be treated as reserved for the humanities in the spirit of Diltheyan distinction between explaining (in sciences) and understanding (in the humanities).

Keywords: Communication, humanities, infrastructure, monograph

BBT User stories. Benefits from joining the thesaurus federation

Helen Goulis *¹, Eleni Tsoulouha *²

¹ Academy of Athens – Greece

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

The Backbone Thesaurus (BBT for short) [1] is the research outcome of work undertaken by the Thesaurus Maintenance WG in an effort to design and establish a coherent overarching meta-thesaurus for the Humanities, under which specialist thesauri and structured vocabularies used across scholarly communities can be integrated and form a thesaurus federation. Its core feature is that it promotes alignment of cutting-edge terminology to the well-formed terms of the meta-thesaurus capturing general meanings. The BBT favors a loose integration of multiple thesauri, by offering a small set of top-level concepts (facets and hierarchies) for specialist thesauri terms to map to. This way, it enables cross-disciplinary resource discovery, while ensuring compatibility with thesauri that cover highly specific scientific domains and areas of knowledge in development.

The labels and the definitions for the facets and hierarchies of the BBT are available in four (4) languages –namely, English, French, German, and Greek.

The controlled vocabularies/thesauri the concepts of which have been mapped to the BBT to this date are the DAI Thesaurus [2], the DYAS Humanities Thesaurus [3] and the Parthenos Vocabularies [4]. At the same time, members of the working group are working towards integrating the Language of Bindings Thesaurus [5], PACTOLS [6], the Taxonomy of Digital Research Activities in the Humanities (TADiRAH) [7] and the Arts and Architecture Thesaurus [7] with the BBT, at least partially.

The BBT is systematically curated by a cross disciplinary team of editors coming from organizations participating in the TMWG (AA, DAI, FORTH, FRANTIQC/CNRS), through BBTalk [8], an online service designed to support collaborative, interdisciplinary development and extension of thesauri. The BBT versions are regularly uploaded to the DARIAH-EU vocabularies along with all their connections to local thesauri via the ACDH-ÖAW service [9].

Goal of this presentation is to showcase the benefits of joining the BBT federation and become part a gradually growing community of thesauri maintainers. The Thesaurus Maintenance WG will bring forward in a YouTube video linked on a poster via a QR code, designed for the Marketplace of the 2020 Annual Event, the cases of DAI and FRANTIQC. Their experience will help outline the challenges encountered in the curatorial process and how they were solved. We

*Speaker

envisage that this presentation will serve as an overall evaluation of the work performed thus far by the Thesaurus Maintenance WG.

References

<https://www.backbonethesaurus.eu>

<http://thesauri.dainst.org/de/labels/f04f41ba.html>

<https://humanitiesthesaurus.academyofathens.gr/>

https://isl.ics.forth.gr/parthenos_vocabularies/

<https://www.ligatus.org.uk/lob/>

<https://pactols.frantiq.fr/opentheso/>

<http://tadirah.dariah.eu/vocab/>

www.backbonethesaurus.eu/BBTalk

https://vocabs.dariah.eu/backbone_thesaurus/en

Current Main Editors: Martin Doerr, Gerasimos Chrysovitsanos, Helen Goulis, Helen Katsiadakis, Patricia Kalafata, Yorgos Tzedopoulos, Blandine Nouvel, Evelyne Sinigaglia, Camilla Colombi, Lena Vitt, Chrysoula Bekiari, Eleni Tsoulouha, George Bruseker, Lida Harami. DARIAH Backbone Thesaurus (BBT): Definition of a model for sustainable interoperable thesauri maintenance, Version 1.2.2, May 2019 (<https://www.backbonethesaurus.eu/sites/default/files/DARIAHBBT%20v%201.2.2.pdf>)

M. Daskalaki, L. Charami, 2017, *A Back Bone Thesaurus for Digital Humanities*, ERCIM News 111, October 2017, Special theme: Digital Humanities (<https://www.backbonethesaurus.eu/sites/default/files/BBT%20News-111.pdf>)

M. Daskalaki, M. Doerr, 2017, *Philosophical background assumptions in digitized knowledge representation systems*, in *Dia-noesis: A Journal of Philosophy*, 2017, Issue 3, p 17-28. (<https://www.backbonethesaurus.eu/sites/default/files/BBT%20News-116.pdf>)
Christos Georgis, George Bruseker and Eleni Tsoulouha, 2019, *BBTalk: An Online Service for Collaborative and Transparent Thesaurus Curation*, ERCIM News 116, January 2019 (<https://www.backbonethesaurus.eu/sites/default/files/BBT%20News-116%29.pdf>)

Keywords: thesaurus building, thesaurus management, thesaurus integration, controlled vocabularies

Integrating European Literary Bibliographical Data: Open Bibliodata Initiative

Vojtěch Malínek * ¹, Tomasz Umerle *[†] ², Cezary Rosiński 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

The poster aims to present the initiative to create a service that would integrate European bibliographical data in literary studies, starting from following countries: Czech Republic, Hungary, Poland, and Slovakia.

This initiative is led by the Czech and Polish Literary Bibliographies, and its development is one of the first activities of recently established DARIAH-ERIC Working Group "Bibliographical Data".

The poster will address:

- 1) the motivation behind integrating bibliographical data,
- 2) challenges for development,
- 3) guiding principles of the service,
- 4) initiative's roadmap.

The goal of the poster is to gather interest from possible stakeholders (users, or data providers).

Motivation

Subject bibliographies are indispensable resources: they organize knowledge crucial to given discipline. As such they can serve as a reliable digital representation of knowledge production (and as such, for example, shape digitization policies), and become a basis for data-driven studies.

There are robust resources of open bibliodata: from specialized bibliographies with detailed subject description, through general library catalogues published under free licences, to metadata from scholarly repositories. Bibliodata researchers, and programmers can process them easily.

*Speaker

[†]Corresponding author: tomasz.umerle@ibl.waw.pl

For literary studies bibliodata is fundamental: bibliodata serves not only as a source of information for further research, but it allows to perform data-driven studies on literary communication systems, book history, etc.

Sustainability: undoubtedly, integration of digital services in the humanities contributes to better sustainability, and systematically reduces interoperability issues for data-driven studies.

Main challenges

Understanding users' needs: in what way international bibliodata service might help scholarly community, especially when the cultural output is created in vernacular languages?

Data extraction, and processing: processing different metadata formats or even national/institution specific interpretations of general standards (e. g. MARC21), translating scope of bibliography into data extracting procedures, also from general bibliographies

Metadata alignment: aligning multilingual vocabularies, and different classification methodologies applied to create different bibliographical resources

Guiding principles:

1. Clearly defining, communicating, and documenting the scope and structure of published metadata,
2. Consistently providing data of both scholarly, and artistic nature - from books, journals, grey literature, and web resources, and different document types - articles, books, parts of the books, online content, etc.
3. High-quality and interoperable metadata that allows both for inter-cultural, international queries, and is consistent with FAIR principles, especially: crucial aspects of metadata are aligned internationally recognized controlled vocabularies (LCSH, LCGFT, COAR, VIAF, Geonames, etc.)
4. Community outreach: reaching out to the scholarly community, but also all those interested in literature by explaining the advantages of bibliodata integration.

Roadmap

First step:

Integration of Czech and Polish - both scholarly, and artistic - literary bibliodata:

0,7 mln records from Polish Literary Bibliography, from 1989 to 2008, covering Polish literature and reception of foreign literatures in Poland,

0,6 mln records from Czech Literary Bibliography, from 1945 to 2020, covering Czech literature and its reception in the Czech Republic.

Next step:

Scaling-up of the service - acquisitions of:

data automatically retrieved from Polish printed bibliographical publications (1939-1988),

data from the Retrospective Bibliography of Czech Literature (1770-1945)

Hungarian, Slovakian literary bibliodata,

CC0 bibliodata from European sources.

Keywords: literary bibliographies, databases, interoperability, open data

DARIAH-EU "Bibliographical Data" Working Group 2020

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

Idea:

The proposed poster will present activities of the "Bibliographical Data" Working Group. The WG was established within the DARIAH consortium in 2019 and operates under VCC3: Scholarly Content Management. The main goal of the Working Group is to serve as a platform for knowledge exchange, and foster cooperation between those who work with bibliographical data: data creators, scholars interested in data-driven research, theorists of bibliography and documentation, IT specialists and other related stakeholders.

Bibliographical Data Working Group focuses on the various aspects of bibliographical data life cycle and covers the wide scope of the themes related to bibliodata. In particular, the Working Group activities concentrate on the following issues:

Data curation

Data quality management

Data preparation for advanced research

Facilitating international data-based cooperation

Methodological issues

Publishing bibliographies

Processing bibliographical metadata

Remediation of bibliographical information

Development of user-oriented services

*Speaker

†Corresponding author: malinek@ucl.cas.cz

Activities in 2019:

After the preparatory phase, the first kick-off meeting with cca 20 attendees was organised during DARIAH Annual Event in May 2019 in Warsaw. Subsequently, the official approval for Working Group establishment was sent to DARIAH Management in July 2019. The second in-person meeting took place during DH2019 conference in Utrecht.

As the main task for the WG in 2019 and 2020 WG members have identified the preparation of the bibliographical data landscape analysis in the humanities.

Working group was formally accepted by the DARIAH management in September 2019.

Currently the Bibliographical Data Working Group consists of around 30 members from cca 15 European countries. Its activities can be followed mainly via the WG Twitter account (<https://twitter.com/bibliodataWG>).

Plans for 2020:

The main task for the WG is the preparation of the landscape analysis of bibliodata in the humanities.

The analysis will be centered around the following topics: bibliodata curation, bibliodata research, bibliodata tools, and bibliodata collaboration. Each of these parts will briefly analyse different aspects of bibliodata landscape, and the conclusions shall facilitate cooperation between scholars working with bibliodata, and become a basis for the future WG activities. We would love to use the occasion of the DARIAH Annual Event in Zagreb to present first results of our analysis.

Other planned activities:

Developing relations with other DARIAH-ERIC WGs,

Fostering opportunities for bilateral or multilateral projects of the WG members,

Working towards integration of literary bibliodata (special poster is planned for DARIAH Event 2020)

Promotion of the WG and bibliodata in academia via conference presentations, specialized workshops, tutorials, etc. (special workshop is planned for DARIAH Event 2020),

Obtaining funding for the WG projects.

Keywords: bibliographies, bibliographical data, metadata, working groups

Introducing the AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) multilingual research tool

Róbert Péter ^{*†} ¹, Zsolt Szántó ¹, József Seres ¹, Vilmos Bilicki ¹, Gábor Berend ¹

¹ University of Szeged [Szeged] – Hungary

The objective of this demo is to demonstrate the workflow, different analytical functions and features of the multilingual AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) data-driven digital tool, which has been developed by NLP researchers and a digital humanist since 2017. This web application enables digital humanists to *critically* analyse the bibliographic data and texts of large corpora and digital collections at scale. The unique features of the AVOBMAT toolkit are that (i) it combines state-of-the-art bibliographic data and computational text analysis research methods in one integrated, interactive and user-friendly web application; (ii) the implemented analytical and visualization tools provide interactive close and distant reading of texts and bibliographic data. It allows users to filter the uploaded datasets by metadata and full-text searches of various types and perform the bibliographic, network and NLP analyses on the filtered datasets.

In the preprocessing phase the user can set nine optional parameters such as lemmatization (spaCy, LemmaGen), context and stopword filtering. Users can create different configurations for the different analyses and visualizations. The metadata enrichment includes the automatic identification of the gender of the authors and automatic language detection of the documents.

Empowered by the Elasticsearch engine users can search and filter the uploaded and enriched bibliographic data and preprocessed texts in faceted, advanced and command line modes. AVOBMAT also supports fuzzy, proximity searches and regular expression queries.

Having filtered the uploaded databases and selected the metadata field(s) (107 in number), users can (i) analyze and visualize the bibliographic data chronologically in line and area charts in normalized and aggregated formats; (ii) create an interactive network analysis of maximum three (meta)data fields; (iii) make pie, horizontal and vertical bar charts of the bibliographic data of their choice according to the provided parameters.

As for the content analysis, the diachronic analysis of texts is supported by the N-gram viewer. The n-grams with a maximum 5-word length are generated at the preprocessing stage, along with the calculations of eight different lexical richness metrics.. Two special types of word cloud analyses are implemented: the significant text (Elasticsearch) cloud showing what differentiates a subset of the documents from others and the TagSpheres (Jänicke and Scheuermann, 2017)

*Speaker

†Corresponding author: robert.peter@ieas-szeged.hu

enabling users to investigate the context of a word. There are bar chart versions of the different word clouds that present the applied scores and frequencies. The traditional close reading examination is fostered by the TagSpheres and the Keyword in Context representations of the search queries.

AVOBMAT has an in-browser Latent Dirichlet Allocation function to calculate and visualize the topic models (e.g. time series, topic correlations). Besides setting the number of iterations and topics, it allows for the adjustment of the LDA alpha and beta hyperparameters.

The export functions of AVOBMAT facilitate the reproducibility of the results and transparency of the preprocessing and text analysis. It helps users realize the epistemological challenges, limitations and strengths of computational text analysis and visual representation of digital texts and datasets. AVOBMAT will be released in the second half of 2020.

Keywords: digital tool, bibliographic data analysis, distant reading, close reading, topic modeling, ngram, lexical richness, network analysis, Elasticsearch, spaCy

Project Znameniti: spatial humanities and deep mapping

Nevenka Kalebic * ¹, Vlatka Lemić *[†] ²

¹ University of Zadar – Croatia

² University of Zagreb – Croatia

The construction of the portal Znameniti.hr started in 2016 with the aim of gathering and unified search of digital material of the leading figures of Croatian culture, science and art from various collections / repositories of Croatian cultural, artistic and scientific institutions. The purpose of the project was to develop a collaborative platform that would make non-commercial scientific, cultural and artistic digital content more accessible to researchers and the general public. Over time, partner institutions have established a collaborative model for portal development and maintenance, and through a variety of projects, work is being done to increase the scope and diversity of digital content represented, to develop new functionalities and to include various institutions that have valuable contributions in their digital collections / repositories about prominent figures of the Croatian past and present into the project.

The Znameniti Project is trying to keep up with the latest developments in the field of new technologies, especially in the field of digital humanities and trying to apply the latest technological advances that will make it more attractive to scientists and the wider user population. With the development of new technologies, especially GIS (Geographical Information System) space has become an important concept of the humanities. The application of GIS in the field of the humanities not only achieves the ordinary spatial mapping of geographical space, but a better understanding of the concept of place as a space with its own history and memory. This use of new technologies within the humanities leads to a new multidimensional and multidisciplinary approach called spatial humanities, whose main research focus has been deep mapping.

Spatial humanities, as a field of digital humanities, develops the assumption that the development of new technologies will allow multidimensional representation of space.. Such processes are called deep mapping precisely because of the possibility of multiperspectival exploration as well as deep connectivity at different spatial and temporal scales.

This paper focuses on prominent writers and the place / institution where digital content about them is represented. The aim of the paper is to give an account of the deep mapping of the place / institution where individual writers are stored. The mapping will attempt to cover all significant historical, cultural and social aspects that have influenced the development of individual institutions. The purpose is to raise awareness of the importance of the institution as a place where memory is being preserved, as well as to contribute to a better interpretation of the lives and activities of famous writers. This approach will try to determine in more detail in which places / institutions the digital content of individual writers is most represented, as well as the possible factors that may have influenced their representation in that institution. With

*Speaker

[†]Corresponding author:vlatka.lemic@unizg.hr

deep mapping, place takes on a whole new dimension in the virtual environment and with all its significant historical, cultural and social characteristics, it greatly facilitates the availability of a wider range of information. With the deep mapping approach, the place takes an important role in the humanities as a source for various scientific research.

Keywords: Project Znameniti, digital humanities, spatial humanities, deep mapping

Insights on scholarly primitives from Digital Humanities research in Spain

Maurizio Toscano ^{*†} ¹, Aroa Rabadán Reyes ², Salvador Ros Muñoz ^{‡3},

Elena González-Blanco ³

¹ Universidad de Granada – Spain

² Universidad Complutense de Madrid – Spain

³ Universidad Nacional de Educación a Distancia – Spain

Digital Humanities are experiencing a growing interest in Spain, especially in the last decade, becoming a leading trend in research, either as a field of study or as a preferential financing topic. At the same time, because of their novelty, they are under scrutiny by the research community and government institutions because the return of investment is not understood neither the role that Spanish researchers can play within European-wide research infrastructures, such as DARIAH.

In order to provide the global community of scholars working in this field with a greater understanding of the current Spanish scenario, LINHD has recently promoted an ongoing research on the evolution of Digital Humanities in Spain in the last 25 years, a timeframe comparable with Unsworth first formulation of scholarly primitives. The immediate goals of the study were to identify researchers in the field of Digital Humanities and to explore their financing, institutional affiliations, research projects and developed resources. The research has been very much data oriented and quantitative at its core, in order to quantify and describe initiatives, researchers, projects, digital resources, educational courses and scientific publications, connected among themselves and spanning from the nineties to the latest contributions. From a quantitative point of view, we collected bibliographical records from over 400 authors, that represent a good approximation of the available literature produced by researchers affiliated with Spanish institutions and a quantifiable measure of the impact and interest in DH within the Spanish research community in the Humanities. More than 360 projects have been mapped, generally small from the point of view of economic resources, but that together represent a significant amount of research funds dedicated, in the last twenty years and by a variety of public and private funding bodies, to research in this field (over 20 million euros). Finally, a dozen educational courses and over 80 digital resources of diverse nature (repositories of documents, collections of artefacts, crowdsourcing platforms, dictionaries, ...) have been analyzed, the latter, most of the time, produced with the aim to publish a service to improve the basic of day-to-day research in the Humanities.

In the context of this contribution, we plan to explore and exploit this relatively vast amount of data in order to identify how the introduction of digital tools and methods from Computer Science has affected the basic functions of research in the Humanities in Spain. Among the types of records collected, we believe that digital resources in particular will be able to provide

*Speaker

†Corresponding author: maurizio.toscano@gmail.com

insights, because they speak more about the reality of research workflows. To perform this analysis, resources will be classified and described according to classical scholarly primitives (discovering, annotating, comparing, referring, sampling, illustrating and representing), in order to highlight presences, absence and recurring associations of these categories, at certain specific stages and over time. Additionally, being resources already classified by discipline (Philology, History, Archaeology, History of Arts, ...) and typology, we will be able to visualize the relationships between scholarly primitives and other dimensions in our data. This taxonomy exercise will also provide an opportunity to reflect about the need for new possible classifications, and on how classical primitives can assume different meanings depending on the scope of a project.

Keywords: digital resources, scholarly primitives, Spain

OPERAS services - Metrics Service, Certification Service, Publication Service Portal and Discoverability Service

Yoann Moranville * ¹

¹ DARIAH – DARIAH-EU – France

OPERAS is a Research Infrastructure supporting open scholarly communication in the social sciences and humanities (SSH) in the European Research Area. Its mission is to coordinate and to federate resources in Europe to efficiently address the scholarly communication needs of European researchers in the field of SSH. OPERAS aim is to make Open Science a reality for research in the SSH. We want to achieve a scholarly communication system where knowledge produced in the SSH benefits researchers, academics, and students fully and widely, and where it also benefits the whole society across Europe and worldwide.

In doing so, OPERAS, also thanks to its services, aims at allowing use of the scholarly primitives, including discoverability, annotating or even comparing. Of course, the focus of OPERAS being on scholarly communication, not all scholarly primitives are planned to be covered here.

The discoverability service, being developed within the TRIPLE project and funded by the European Commission, will therefore serve as the discoverability platform of OPERAS and will be available to researchers in order to discover 3 main types of resources: data(sets), researchers' profiles, and projects – data here is mostly meant as publication data. By aligning the metadata retrieved from various sources (either harvested or retrieved via APIs) and enriching it (also with the help of machine learning), comparison will also be possible on the metadata level of the resources.

Also used not only by the researchers, but also useful for publishers and funders, the Metrics Service developed in the HIRMEOS project and now in beta, currently linked to 5 different partner platforms allows usage metrics to be retrieved by a central service and available to all via its API. These Metrics are usage data coming from various sources such as views or downloads on multiple platforms (i.e. Google Books, World Reader, etc...) as well as Altmetrics coming from mentions on other platforms such as Twitter, or the annotation platform hypothes.is.

A Publishing Service Portal is currently being implemented as part of the OPERAS-P project, also funded by the European Commission. This portal, which is partly based on an already developed prototype made by DARIAH within its Humanities at Scale project, will provide to SSH researchers a simple and single entry-point in order to reach the different publishing services provided by OPERAS and its members.

OPERAS Certification service, provided by the DOAB (Directory of Open Access Books) plat-

*Speaker

form and developed within the HIRMEOS project, and already operational in beta, is intended to certify publishers at publishers' and at single publications' levels. The peer-review information is then available via a public API as well.

There are other services also available or being developed, but the focus will be done on those 4 developed or soon to be developed services. We will also present how the services will be brought and available to a larger community by reaching out when they are successfully available via the EOSC Marketplace.

Keywords: OPERAS, services, infrastructure, open science, discovery, metrics

Echo from the War: Preserving the Memory of the Bulgarian Veterans from WWII

Konstantin Golev * ¹

¹ Institute for Historical Studies at the Bulgarian Academy of Sciences – Bulgaria

The project *Echo from the War* started in 2008 by two graduate students (Konstantin Golev and Marian Gyaurski) with the aim to collect oral history interviews Bulgarian veterans from WWII. In the years to follow the project expanded significantly and the team included many new members. By acquiring funding from the Center for Excellence in the Humanities "Alma Mater" the team managed to conduct a national-wide field work to gather oral history data from all the 28 administrative regions of Bulgaria in the course of a three years-long field work campaign. Thus, a collection of about 400 separate oral history interviews covering all regions of Bulgaria was formed. The collection includes representatives of almost all types of troops as well as members of almost all ethnic and confessional minorities of Bulgaria. A number of female veterans was also interviewed. In addition, the collection includes respondents with different social status as well as with various political affiliation. Thus, the members of the team tried to reflect the political diversity in Bulgaria during the war and to trace the influence of the opposing political views on the interpretation of key events from the period. Contrast were sought in particular among the members of the Communist party who made rapid social advancement after the war and members of opposition groups and parties (such as BZNS, BNL, etc.) who were often repressed in the years after the conflict.

During the last decade the team was concentrated on the field work given the advanced age of the respondents (currently about 96-97 years old) and the rapid decline in their numbers. Yet 39 selected interviews have been published in two consecutive volumes (2011 and 2015) and have been made accessible on the website of the research group: <https://veterani.uni-sofia.bg/interviews>; where the two volumes are also available.

Now the team faces the challenge to extract and interpret the information of a field work material that have been gathered for more than a decade. Such a huge quantity of data suggests the usage of Digital Humanities and the first steps towards the digitalization process have been made during the fieldwork. Thus, all the interviews are conducted and preserved in a digital form (mainly audio files, many pictures of the respondents, a number of photos from the period of the war, as well as few video records). The resulting database has been organized according to geographical principles and the interviews have been ordered according to the modern administrative division of Bulgaria, reflecting the place where the respondent lived at the moment of the interview. Yet the real challenge is how to apply the achievements of the Digital Humanities in the most time-consuming processes such as the transcription of the interviews as well as how to use the modern technologies for parallel analysis of the data, drawing of a social network of the respondents and the persons they mention, as well as mapping the individual fighting route of each veterans. The aim of the present paper is to present what has been achieved so far and to look for solutions of the technical problems faced by the team.

*Speaker

Keywords: Oral history, World War II, Interviews, Veterans

Open Badge Ecosystem for the Recognition of skills in Research Data management and sharing

Tomasz Umerle * ¹

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

The poster aims to present the Open Badge Ecosystem for the Recognition of skills in Research Data management and sharing [OBERRED] project, and garner interest from potential stakeholders.

Project is funded by Erasmus+ programme (Cooperation for innovation and the exchange of good practices), and realized by the consortium of 7 European institutions, led by the University of Nice Sophia Antipolis.

Main goals

OBERRED project combines Open Sciences and Open Education to facilitate skills accreditation in research data management.

There are two main objectives of the project:

1) OBERRED will prepare a set of interconnected practical guides to support the community of facilitators of RDM ecosystem who are willing to support the ecosystem with Open Badges concepts, and methodology:

- Basics of Open Badges,
- Basics of Research Data Management,
- How to facilitate RDM processes with Open Badges?

2) OBERRED will create a collaborative network, a small ecosystem of Open Badges in Open Science at European level to acknowledge acquired competences in sharing and managing research data in the framework of the Open Science:

- RDM skills will be analyzed thanks to interviews conducted with RDM practitioners,

*Speaker

- the recognized skills will be mapped to ESCO framework of competences
- a set of Open Badges will be issued to RDM stakeholders.

Innovating the ecosystem - the role of facilitators

The crucial part of the project will be the promotion the OB use in RDM communities by appointing a group of facilitators of the RDM ecosystem who would work with different RDM stakeholders on Open Badges implementation in RDM skills accreditation.

Open Badges prepared within the project will reflect the actual skills used, and recognized by stakeholders.

Facilitators will work closely with stakeholders - provide training, and support - to take advantage of Open Badges concepts, and help facilitate RDM ecosystems.

Timeline

The project started in September 2019, and will be finalized in August 2022.

Project will start from the development of the online trainings (2020), and practical guide (2021), and proceed to RDM stakeholders analysis, and cooperation (2021-2022).

Partners

Universite de Nice Sophia Antipolis (France; leader)

Beuth-Hoschschule für Technik (Germany)

Centre National de la Recherche Scientifique (France)

FernUniversität in Hagen (Germany)

Instytut Badań Literackich Polskiej Akademii Nauk (Poland)

Università degli Studi di Napoli Federico II (Italy)

World University Services of the Mediterranean-WUSMED (Spain)

Contacts

Project manager: Nicolas Hochet (nicolas.hochet@univ-cotedazur.fr)

Poster prepared by: Tomasz Umerle (tomasz.umerle@ibl.waw.pl)

Keywords: research data management, RDM, skills recognition, open science, open education, open badges

Retroconversion of bibliographical publications: main stakes and challenges

Patryk Hubar ^{*†} 1, Nikodem Wołczuk 1

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

The main purpose of this poster is to present the importance of retroconversion of bibliographical documents which are crucial in data-driven research activities in areas such as literary research, cultural and art studies.

What is retroconversion of bibliographical publications?

Retroconversion understood as transformation process of printed bibliographical publications into human- and machine-readable database can be also enriched with such elements as linked open data and semantic web mechanisms which can create an unique, modern and rich source of information about knowledge production in various research domains, which translates both into facilitating scientific research and also into creation of new research areas and topics.

Main goals and challenges

In particular, the poster will present the following challenges/goals:

- challenges of transforming varied bibliographical elements from different sources and periods into one, structured and consistent database,
- advantages of creating open and structured bibliographical database,
- integrating bibliographical metadata with other, similar solutions in different research fields and countries,
- benefits of enriching converted metadata by additional information and links to external sources.

The main purpose of the poster is to draw the attention of scientists and academic and research units to the issue of access to bibliographic data.

Research project info

The poster will be based on findings of PhD research project Retroconversion of Polish Literary Bibliography which is being prepared at the Institute of Literary Research of the Polish Academy of Sciences (IBL PAN) and Polish-Japanese Academy of Information

*Speaker

†Corresponding author: patryk.hubar@ibl.waw.pl

Technology (Digital Humanities postgraduate studies), in cooperation with Department of Current Bibliography of the IBL PAN.

Project is focused on transforming Polish Literary Bibliography (<https://pbl.ibl.waw.pl/>) - which is one of the most important sources of information about Polish literature and culture in Poland - into structured and linked database using a number of algorithms and machine learning solutions.

Project of retroconversion of Polish Literary Bibliography is also supported by DARIAH-ERIC Working Group “Bibliographical Data”.

Contacts:

Patryk Hubar, Institute of Literary Research of the Polish Academy of Sciences, Poland, patryk.hubar@ibl.waw.pl

Nikodem Wolczuk, Department of Current Bibliography, Institute of Literary Research of the Polish Academy of Sciences, Poland, nikodem.wolczuk@ibl.waw.pl

Keywords: literary bibliography, metadata, semantic web, data analysis, linked open data

Towards a Global Study of Algorithms thanks to Digital Humanities and Sociology

Jeremie Poiroux * ¹

¹ Centre Marc Bloch – Germany

Over the last few years, a research object has been attracting the attention of quite a number of members of the scientific community of digital humanities and sociology: the study of algorithms, and more precisely of their impact on society. Examples include the study of the impact of work distribution algorithms (Uber), recidivism prediction algorithms (COMPAS) or recommendation algorithms (Youtube or Spotify). Both quantitative and qualitative methods are used for these studies. At the same time, we can find an increasing integration of the digital humanities in these studies, both to analyse the data used by the algorithms and their results. To a more limited extent, simulations of the operation of these algorithms can be found. However, and this is the purpose of this poster, an insufficient number of studies in this field have focused on the fabrication of these algorithms, either on the way engineers and more widely organizations design these programs that have the impacts that we now know a little better (their non-neutrality or discrimination to name only two).

Sociologists can study this fabrication by using qualitative methods by interviewing designers or by conducting immersion studies in the companies which develop these algorithms. Quantitative methods are, to our knowledge, not very relevant (the chances of accessing a critical mass of data are low). On their side, the digital humanities have initially brought methods or reverse engineering solutions, i.e. attempts to open the black box and understand how the algorithm works technically.

As co-author of a sociological study being finalized on the design of recommendation algorithms based on thirty interviews, I would like to present the following two challenges with a poster exhibited at the DARIAH Annual Event 2020. The first stake is to give a new perspective to the work on the ethical issues raised by algorithms. The analysis of the responses of the 30 engineers provides a rich portrait of their apprehension of work, particularly through the dichotomy between “high ideals” (such as educating citizens) and “low ideals” (such as optimising a particular variable).

The second major challenge is to propose ways in which the digital humanities and sociology could collaborate to study the making of algorithms that would go beyond the use of tools to analyze and categorize interview material. The whole point here is to launch a reflection to find a relevant and innovative method allowing the global study of algorithms, from their (theoretical) conception to their impacts, going through their fabrication and their use, in order to expand the production of knowledge. The process of analyzing the interviews on the recommendation algorithms design and the study’s writing should allow to bring the first elements of answer, to open the discussion and to revisit the seven scholarly primitives.

*Speaker

Keywords: algorithms, ethics, interdisciplinarity

Digging for unspecified information requirements: a case study of Digital Library of Arne Novák users

Michal Lorenz ^{* 1}, Pavla Martinkova ^{*† 1}

¹ Masaryk University, Faculty of Arts [Brno] – Czech Republic

Surveying the information needs of users is already a key issue in the preparatory phase of the research infrastructures. In this phase the integration of research platforms is being prepared. A research of information requirements of digital scholars is not limited only by the difficulty of reaching users, but also because of the lack of transparent platform architecture, log data cannot be used effectively. Users of platforms are often happy to get access to data and resources relevant to their research. They consider most of their work to be abstract, intellectual, or simply of such nature that computer and information technology cannot help them. Thus, many needs are not transformed into requirements, nor are they perceived as needs. In order to identify the unconscious demands of users and to find the potential for improving the research platform, we decided to make a sense-making analysis of six researchers who use the Digital library of Arne Novák in their research. Digital library of Arne Novák is a research platform providing access to the work of a prominent figure in Czech literary science, that we want to make available to the LINDAT/CLARIAH-CZ research infrastructure. A micro time-line interview was conducted with all researchers to understand how researchers use research platform, how they interact and practice research with digital data and how we can shape a vision of augmentation of human intellectual possibilities when working with this platform. The research verified that in this way we can identify unnamed and unspecified information requirements of research data users.

Keywords: Information requirements, research information practices, research platforms, LINDAT/CLARIAH, CZ, Digital Library of Arne Novák, Sense making analysis

*Speaker

†Corresponding author:martinkova@phil.muni.cz

DARIAH public consultation

The SSH Open Marketplace – Involvement of the research community in the development process

Laure Barbot *¹, Frank Fischer *², Tomasz Parkola *³,

Stefan Buddenbohm *[†] ⁴

¹ Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

² Digital Research Infrastructure for the Arts and Humanities – Germany

³ Poznan Supercomputing and Networking Center – Poland

⁴ Göttingen State and University Library – Germany

The SSHOC cluster project (Social Sciences and Humanities Open Cloud) is a contribution to the European Open Science Cloud (EOSC) and addresses particularly the requirements of the Social Science and Humanities research community. An important goal of the project is to deliver the SSH Open Marketplace (<https://sshopencloud.eu/marketplace>) as a discovery portal and a comprehensive information hub for the SSH research community regarding tools, services, training materials and other resources. Beyond this straightforward approach of identifying resources the marketplace also utilises the research process as an underlying matrix, e.g. in addressing research methods as a composition of actions linked with resources, and where possible, connect them with tutorials, feedback or assessment functions.

This sketch of the SSH Open Marketplace obviously connects very well with the scholarly primitives as described by John Unsworth – discovering, annotating, comparing, referring, sampling, illustrating, and representing – as they shape the requirements the Marketplace has to meet to be of actual value for the research community. This also aligns with the DARIAH Strategic Plan 2021–26 which stands upon four pillars: (1) building the Marketplace as a humanist-friendly component of the European Open Science Cloud, (2) approaching training and education strategically and in a coordinated fashion, (3) deepening our connection to our communities and ensuring they are with us, and (4) strengthening our voice in policy and advocacy.

Our workshop will encourage and enable important feedback from the community for the refinement of the concept and development of the marketplace one month before its alpha release in June 2020. A testing instance of the marketplace is available here: <https://sshoc-marketplace.acdh-dev.oeaw.ac.at/> SSHOC follows an agile development approach with user feedback from the very beginning and feedback loops in between. The audience of the DARIAH Annual Event – humanities scholars and the DARIAH community – is a perfect multiplier for

*Speaker

[†]Corresponding author: stefan.buddenbohm@sub.uni-goettingen.de

us to receive feedback from the research community. Beyond this we invite the DARIAH community at large, members of the DARIAH working groups, NCCs, students and researchers.

The event is intended as feedback loop to present and discuss the actual state of affairs of the SSH Open Marketplace and is one in a series of events implementing a participatory design approach. To this end, we will have a thorough look at the platform (including an interface walkthrough), its functions and its contents and try to map this to individual research practices. In short: Does the SSH Open Marketplace meet the requirements of the DARIAH communities from the Arts and Humanities SSH community in terms of content and functions and if not entirely, what is missing? This question particularly addresses the sufficiency of the Marketplace in terms of its content and usability.

The half-day workshop on May 27th is intended as interactive activity and hands-on learning opportunity. It will be organised by the SSHOC Work Package 7. This may include audio/video recordings on location.

Keywords: SSHOC, Marketplace, DARIAH, SSH Open Marketplace, agile development, feedback, user requirements, community

Scholarly Primitives in Theory and Practice

Digital Humanities or humanities in digital?

André Pacheco ^{*†} ¹

¹ University of Coimbra – Portugal

Over the past few decades we have been witnessing the availability of an array of novel digital tools and technologies for data creation, collection, analysis and dissemination that have permeated and transformed scholarly practices across all disciplines. Curiously, in the humanities, unlike in any other field of science, the use of these new tools has given birth to a new, still divisive movement called Digital Humanities. As a result, the boundaries between humanities and digital humanities has become a grey zone that requires debate which I believe that should focus not on what changed, but rather what remains at the core of scholarly practice. As such, this paper aims to provide a much necessary theoretical contribution to the discussion of the nature of this field.

This study uses a mixed methodology that combines qualitative methods of source selection and documentary research with quantitative methods of data analysis. The initial sample is comprised of articles indexed in the core collection of Web of Knowledge, under the topic "digital humanities", and the "Information Science, Library Science" category. Then, a quantitative analysis of the yearly publications, regions, publishing languages is undertaken in order to contextualize scholarly production in Digital Humanities, from this database's perspective. The discussion is initiated with notion of scholarly primitives (discovering, annotating, comparing, referring, sampling, illustrating, representing), as formulated initially by John Unsworth in 2000. These are compared to the suggestion of Palmer and colleagues, in 2009, of scholarly information practices, and to Project Bamboo's 2010 themes of scholarly practice. These primitives aim to be staples across era and media for scholar activities, and act as checklists for building new digital tools. Therefore, they are suitable as a starting point for the discussion of scholarly practices, specifically, the practices of digital humanities scholars, with the aim of contrasting them to oldschool practices of humanists. The discussion is then fueled by articles retrieved through the qualitative source selection. The evolution of the term 'Humanities Computing', as initially appeared in 2004 in Blackwell's *A Companion to Digital Humanities*, towards 'Digital Humanities' is analyzed in order to illustrate changes in scholar mentality, identity and culture. As a conclusion, it is argued that the 'who', 'what' and 'why' of research in the humanities remains immutable, but the 'how' is transformed by digital tools. The practices of scholarly primitives might change, with the diluting of secular borders between sciences and the humanities, but the research principles do not. Also, it is noted that Digital Humanities, more than a well-defined field, should be understood as a community characterized by a culture of collaboration, networking and openness to novel research tools that allow for unprecedented methods for creating, processing and sharing research data.

*Speaker

†Corresponding author: andrez.pacheco@gmail.com

Keywords: Digital humanities, Web of Science, Library and Information Science, scholarly primitives

Scholarly primitives from a workflow perspective

Marijn Koolen ^{*† 1}, Liliana Melgar ², Sanna Kumpulainen ^{*3}

¹ KNAW Humanities Cluster – Netherlands

² Universiteit Utrecht – Netherlands

³ University of Tampere [Finland] – Finland

The concept of "scholarly primitive" has been widely welcomed both by humanists and system designers in the humanities, due to the fact that it made it possible to have a solid conceptual basis for the operationalization of the essential functionalities required for advancing computer-mediated work in the humanities. It has also helped to prioritize, and to provide general frameworks for the analysis of system requirements, which otherwise would have remained vague, or too specifically tied to the particular projects.

However, in the design of actual systems or digital infrastructures, most initiatives that have tried to apply the concept of "scholarly primitive" have acknowledged the importance of looking at the relationship between them (e.g., Palmer et al., 2009), but haven't framed these primitives or activities within the larger workflows in which researchers transition from one activity to the other, or perform multiple connected activities in tandem.

One of the most fully-fledged conceptual models for scholarly research activity based on scholarly primitives was developed by DARIAH (note 1). Grounded in conceptual and empirical research, this conceptual model aims to fit the needs of the actual research life-cycle (Benardou, 2013), with the resulting NeDiMAH (note 2) Methods Ontology (NeMO) centered around a list of research activities, which are hierarchically connected by means of a taxonomical structure.

We have developed a process perspective that can help to understand how these activities are interconnected. For example, the activity "gathering" is defined by the NeMO ontology as "aggregating discovered resources [...] for further analysis." (note 3). While this normalization of the activity name and definition is useful for several purposes in system design, the NeMO ontology makes no connection to the activities that occur during "discovering" or "analysis", which makes it difficult to understand how "gathering" occurs in practice, and how it works in real contexts where other activities precede, succeed, or occur simultaneously to the "gathering" activity.

The lack of understanding of the scholarly primitives (or of the research activities) in a workflow perspective has negative implications in designing systems that support the research life-cycle. To overcome this limitation, this paper introduces the concept of "workflow transitions" and presents investigations of scholarly work using this concept. The method we propose consists of

*Speaker

†Corresponding author: marijn.koolen@di.huc.knaw.nl

selecting a sample of representative "research projects" (Koolen et al., 2020) and use them as the unit of analysis for the study of workflows and workflow transitions.

We developed our approach through analysing the research process of two digital humanities projects in great detail, by interviewing the project collaborators (using Critical Incident Technique), studying their papers and looking at datasets, tools and scripts.

In this paper we will present the resulting workflow visualizations, the findings of our analysis and their implications for digital infrastructure support for humanities research.

Footnotes:

<https://www.dariah.eu>

[NeDiMAH.eu](https://www.nedimah.eu)

<http://nemo.dcu.gr/index.php?p=navigate#>

REFERENCES

Benardou, Agiatis, Panos Constantopoulos, en Costis Dallas. "An Approach to Analyzing Working Practices of Research Communities in the Humanities". *International Journal of Humanities and Arts Computing* 7, nr. 1–2 (2013): 105–27. <https://doi.org/10.3366/ijhac.2013.0084>.

Koolen, Kumpulainen, Melgar-Estrada. "A workflow analysis perspective to scholarly research tasks". In *2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*, March 14–18, 2020, Vancouver, BC, Canada. ACM.

Palmer, Carole L., Lauren C. Tefteau, en Carrie M. Pirmann. "Scholarly Information Practices in the Online Environment". Dublin, Ohio: OCLC Research, 2009.
<http://www.oclc.org/research/news/2009/01-23.html>.

Keywords: research workflows, research process, workflow transitions, research methods, digital humanities, information behavior

Beyond scholarly primitives: an epistemological foundations for computational research in the humanities

Fabio Ciotti * ¹

¹ University of Rome Tor Vergata – Italy

The deep and long running influence of John Unsworth's notion of "scholarly primitives" is undeniable. Around the same years W. McCarty and H. Short developed the idea of the "methodological commons", another framework used to map the kind of scholarly activity in the field that was then called Humanities Computing. In many ways it was better at describing the field, but "scholarly primitives" were more successful and had a deeper impact in the self-conceptualization and meta discourse of DH. One may wonder why the idea proposed by Unsworth was so successful. I think that the success of Unsworth's scholarly primitives is rooted in the extreme simplicity of their definition, or better, in the pragmatist framework that governed their formulation. Apart a (somewhat instrumentalist) hint to the notion of *primitive terms* in axiomatic theories as an analogical base for its proposal, both the selection and the description of the scholarly primitives in Unsworth's paper were based on practical examples and anecdotal exemplification. After all, they were developed in the preparation of two project proposals. This pragmatist approach was of course very easy to grab for the community, in as much as it made appeal to the personal experiences of many practitioners of DH. But there is also a deeper reason: it was ideologically and culturally consonant with the phase from the evo of Humanities Computing to that of Digital Humanities. In a sense, it served as base for the rhetoric of scholarly inclusivity that was going to characterize the field and that some years was epitomized by the notorious "Big Tent DH" motto adopted in DH2011 conference at Stanford. The primitives listed by Unsworth were so generic and encompassing that could be used to justify the inclusion inside the field of almost everyone had to with digital resources and/or computational tools, at whatever level of methodological and theoretical involvement. They were a perfect justification for accepting as a valid and sound research in DH both "hard core" stylometry, markup practice, ontology building and "simple" Web sites development. One aspect worth pointing out is that most of the primitives were activities not inherent to the research and new knowledge production activity, but to its infrastructural pre- and post-conditions.

Although I don't want to overlook their historical importance, I think it is necessary to rethink the methodological foundations of the galaxy shaped field of DH in a more epistemologically sound way. My proposal is based on three core layers:

1) the computational methods layer, that pertains to the set of formal methods adopted in research

*Speaker

2) the data model layer, that pertains to the set of data models and data structures used in research

3) the theory/model layer, that pertains to the way theories and models control the research process and gives

Each layer on its part is structured in typologies. For instance, methods are divided in two main sets: fully automatic computing methods vs. human based computing methods. Real and innovative DH research is composed by an interplay of all three layers. The critical (in a Kantian sense) and theoretical study of this methodological framework and of its interaction with humanities theories and scholarship forms the subject of a real disciplinary core of Digital Humanities studies, for which maybe the label Humanities Computing should be taken back to intellectual life.

Keywords: scholarly primitives, methodological commons, methods, methodology of DH, theoretical digital humanities

Reassessing Scholarly Primitives through Experience from Project Factories and Urban Fabric in Bulgaria

Ivaylo Nachev * ¹

¹ Institute for Balkan Studies Center of Thracology – Bulgarian Academy of Sciences – Bulgaria

The proposed paper will look at the concept of scholarly primitives through practical experience gained within the ongoing research project Factories and Urban Fabric in Bulgaria (History and Contemporary Adaptation). Within the project launched in 2018 that combines urban studies, economic history, industrial heritage and digital humanities the research team works on compiling a comprehensive digital archive of varied types of documents. The project focusses on documents that are related to the establishment of modern industry and its long-term impact on urban development in several Bulgarian cities. The main tasks of the project include creating a model for a digital database whose core features will be presented to the audience. In addition to highlighting some results of the project, the paper will elaborate on main challenges related to processing and structuring acquired data. The paper will also present the process of organising the collections and presenting them in a meaningful way through diverse digital channels. Problems of visualisation of parts of the data set will also be discussed. The proposed paper will examine as well opportunities for new digital technologies to accumulate and at the same time to give open and easy access to large data collections. In this particular case attention will be paid to possibilities for utilising such digital humanities projects not only for research purposes but also for raising awareness among the broader public about the value of cultural heritage, including raising awareness of architectural value and possibilities for future uses of deserted factory buildings. Therefore, the project seeks to inter alia promote through digital means opportunities for sustainable research infrastructures.

Keywords: urban studies, industrial heritage, dataset

*Speaker

Scholarly primitives in the field of Glagolitic script: foundations for transition of traditional research activities into virtual research environment (VRE) for Zadar Glagolitics

Marijana Tomić * ¹, Laura Grzunov † ¹, Martina Dragija Ivanović ‡ ¹

¹ University of Zadar. Department of Information Sciences. Zadar, Croatia – Croatia

An interdisciplinary scientific project *Digitization, bibliographic description and research of texts written on Glagolitic, Croatian Cyrillic and Latin scripts until the end of 19th century in Zadar and Šibenik area (further Written heritage)*, is carried out at the University of Zadar in co-operation with Vestigia Manuscript Research Centre of University of Graz, Austria (Tomić, 2018). One of the project goals is to coordinate the interdisciplinary, cross-institutional research of Glagolitic script and culture by providing Virtual Research Environment (VRE), namely digital scholarly platform for research of Glagolitic script of Zadar provenance. The documents involved are multifarious and include epigraphic monuments, historical manuscripts, printed books, art objects, objects of everyday use, and other objects connected with Glagolitic script and culture. The variety of objects of research imply the variety of research disciplines and research methods involved in its research, majority of which are in humanities (linguistics, palaeography, codicology, epigraphy, history, bibliology, heritology, etc.). However, a number of other disciplines are, or should be, involved in research of Croatian Glagolitics in its transfer from traditional to digital research environment, ranging from social sciences to technical sciences (bibliography, information organization, sociology, data bases, data-mining).

In this paper, the methodology of designing VRE which is to support research activities of scholars researching Zadar Glagolitics will be demonstrated. The process of designing VRE is conducted in several phases. The first phase refers to research of usual research practices of scholars researching Zadar Glagolitics. The method used is structured interview with scholars. The practices are mapped to a set of "scholarly primitives" as defined by Blanke and Hedges, namely discovering, collecting, comparing, delivering and collaborating (Blanke, T., & Hedges, M. 2013). The aim of the that phase is to define the set of tasks which have to be transferred to VRE. Its transfer depends on several characteristics, such as the precision of its definition and the possibility of conducting the task in digital environment. Both will be tested with information scientists and practitioners.

The main challenges of the research is to design the VRE enabling interdisciplinary research and capable to address the specific research tasks of specialized scholars, as well as more generic ones. The system has to give the scholars opportunity to use specific digital tools in compliance to

*Speaker

†Corresponding author:lgrzunov@unizd.hr

‡Corresponding author:mdragija@unizd.hr

their research questions and methodologies, but also to bring them a step further – to accelerate their research by the use of digital tools and drive them at using new methodologies to answer the same or generate and answer new questions. Also, the VRE must be designed in such a way to persuade scholars used to traditional methods to use digital tools and research methods in Glagolitic script research within DH framework.

Keywords: Scholarly primitives, Glagolitic script, Zadar, Digital Humanities

3D Scholarly Editions: Scholarly Primitives Reboot

Susan Schreibman * ¹, Costas Papadopoulos

¹ Maastricht University – Netherlands

When John Unsworth conceived of the idea of scholarly primitives, Digital Scholarly Editions (DSEs) were the preeminent form of digital scholarship. Predominantly text-based, with the vast majority using TEI-XML, these editions became exemplars of how scholarly primitives could function in a digital environment: from the Blake Archive's comparing of different textual versions of book plates, to the Rosetti Archive's annotating of Rosetti's oeuvre, to The MacGreevy Archive a platform for discovery of a poet/literary critic's work that was out of print. The goal of this presentation is to explore how well (or even if) the primitives can be used as a theoretical underpinning for scholarly editions in which the text(s) of the edition are represented in three dimensions.

3D DSEs are not inheritors to print-based editions (as first generation DSEs were), but might be thought of as assemblages, little machines of knowledge (Deleuze and Guattari, 1987, 4) in which texts are represented as open objects that can be read and understood through and by their means of production and reception, re-elevating the notion of scholarly primitives within these interlinked discourses which need not privilege the alphanumeric.

In 1999 (around the time Unsworth developed the notion of scholarly primitives) DF McKenzie enumerated the panoply of objects that could be open to the kinds of intensive bibliographical study (transmission, production, and reception) that textual scholars had traditionally reserved for print/manuscript traditions. The pan-glossary that the digital provides allows us to abstract the material medium of these objects to conceptual ones, representing any human activity, not only the artistic but the anthropologic. What is key, however, is not that the object (what we are broadly calling the 'text'), is digitised and distributed electronically, but that it exists within a knowledge site, eg: a Digital Scholarly Edition. Within this framework, we argue for the addition of three-dimensional (3D) (re)constructions of cultural heritage.

This paper will discuss the various forms of 3D representations: from schematic representations of buildings to photorealistic renderings and predictive simulations of ancient structures (see Dawson et al., 2007); and from spatial analysis (see Paliou et al. 2011) and physics simulations (see Oetelaar, 2016) to interactive virtual worlds utilising online platforms (see Sequiera and Morgado, 2013) and game engines (see the projects carried out as part of the Humanities Virtual Worlds Consortium – <http://virtualworlds.etc.ucla.edu/>).

In this paper we also present a framework wherein the 3D models and supporting evidence (textual, social, and historical) are gathered in the form of annotation (Gius and Jacke) 2017

*Speaker

and apparatus (Gabler, 2010, 44) in a knowledge site that encompasses within the same computational paradigm both the primary text and the material that informed the decisions in creating the text, thus providing the community which it serves a tool for ‘prying problems apart and opening up a new space for the extension of learning’ (Apollon et al., 2014, 5-6).

References

upon request

Keywords: 3D reconstruction, digital scholarship, research outputs, publication models, scholarly editions

Tools and Workflows

Unus pro omnibus! Generic research tool for all Humanities disciplines.

André Kilchenmann ^{*† 1}, Flavie Laurens ^{*# 1}

¹ Data and Service Center for the Humanities DaSCH – Switzerland

The "digital turn" has changed research in the Humanities to a large extent: many new digital tools and methods exist with which you can access and analyze texts, videos, sound and music. However, those tools are most of the time standalone applications and it is more difficult to combine various records. A good illustration of this situation is research projects with moving image as main (re)source. Scholars record current events and interview contemporary witnesses like historic or ethnographic projects. Here, moving images or videos need to be transcribed which could be a "simple" interview transcription. But in some disciplines like sociology or film and media studies, these multimedia objects must be extended which complexify the process. In those cases, scholars would also like to annotate the source, to describe the composition of the image, the soundtrack, or the movement of the camera. It's a linkage between various sources and descriptions. The question is: How can we bring them all together?

At the Data and Service Center for the Humanities (hereinafter called DaSCH) in Basel, Switzerland, we have to deal with all different data sets from all disciplines in the Humanities. The DaSCH is a national research infrastructure which provides data handling services like data curation, long-term access, and research and analysis tools to work with qualitative data. We bring a wide variety of data, data models and media (digital representations) from different disciplines together: from archaeology to philosophy; from moving image to books, audio and still images. An important aspect of managing qualitative data in the Digital Humanities is that, in most cases, the preservation of data sets alone makes little sense. We have to store data sets that can be accessed, re-used, connected and annotated.

To reach this goal and to provide qualitative data handling services, the DaSCH develops and maintains a software platform called "Knora" consisting of a database based on a Resource Description Framework (RDF) triple store and Application Programming Interfaces (APIs). Knora handles data from database, as well as media files stored on our own IIIF-based-media server. Those tools are part of the backend, the server side. Scholars with good IT-skills can interact with APIs and work with their data. For scholars with limited IT-knowledge, we need to provide a simple, generic user interface.

We are developing an intuitive, easy-to-use web-based application, called "Knora-App", placed on top of Knora to directly use its powerful data management functionalities. Data models and data will automatically follow accepted standards, be findable, accessible interoperable, and

*Speaker

†Corresponding author:a.kilchenmann@unibas.ch

#Corresponding author:flavie.laurens@unibas.ch

re-usable (FAIR principles). With Knora-App, scholars will have a ready-to-use platform in order to create their own data models, upload data, attach metadata, and perform analyses and data-visualization as they could do with a desktop data management tool. Even scholars with small data sets will have access to long-term accessibility at minimal cost and time to keep their research data alive, guaranteeing longevity of the data.

Keywords: vre, database, rdf, iiif, api, archive, research, data center, web, application

Making the most of digital resources with IIF and OCR enabled transcription toolset

Tomasz Parko-la ^{*† 1}, Micha-l Kozak ^{* 1},

Tomasz Kalota ²

¹ Poznan Supercomputing and Networking Center – Poland

² Wroclaw University Library – Poland

Huge amounts of digital resources are available online - datasets or services are there to be reached/processed. However, one of the challenges in utilizing these resources, is to appropriately address researcher's needs. Although many repositories provide digital assets, their metadata or content is not always relevant for research practices. In the case of scanned/digitised resources, the main challenge pertains to availability of full-text historical documents with uncommon scripts/languages. There have been several approaches to tackle this issue, including manual transcription and automated OCR/HWR techniques (e.g. FromThePage and Transcribus projects). However, there is no approach developed yet to support digital repositories in enhancing their content via transcription tool with built in character recognition, IIF-based streams and automated training routines. In 2019 Wroclaw University Library (WUL) in cooperation with Poznań Supercomputing and Networking Center (PSNC) launched a project to extend their digital infrastructure with such an approach. Since 2011 PSNC has been developing Virtual Transcription Laboratory (VTL) - an open environment for handling scanned textual resources, executing OCR and its post correction, conducting transcription as well as training OCR engine. This new project will add features to VTL or update existing ones, so that the overall solution is able to better respond to the needs of the end users.

The main idea behind the new approach is to fully integrate digital repository with the transcription toolset, so that it is possible to enhance digital assets with full-text. This approach will stimulate cooperation between researchers and content providers - WUL will provide digital content that can be used in the research practices, while the researchers will work with the content and create/enhance its full-text. The important part is that the full-text will be fed back to the original repository as an alternative representation of the digital resource.

VTL is composed of three modules: Import and export, Text recognition as well as Transcription and annotation. Import and export interacts with external systems and deals with various formats. It can import data from IIF manifest, TEI P5 and METS. As a result all references to content files are either to the original source (IIF stream) or to the internal IIF streams which original files were converted to. Text recognition module can then be used to automatically recognize textual representation of the imported images. It is based on Tesseract 4.0 that utilizes neural net (LSTM) algorithm. Transcription and annotation module is the place

*Speaker

†Corresponding author:tparkola@man.poznan.pl

where users can transcribe documents manually or correct Tesseract's recognition results. By default transcription is done on a line level, but this can be changed by the user if necessary, e.g. by adding new regions or removing/replacing them. There are multiple ways to annotate the text itself, e.g. with headers, page numbers, font style or comments. Once the transcription is ready (even partially) the reviewers can verify if the transcription is correct. They can correct errors and provide feedback to the transcriber. Finally, once the transcription is verified the system can export full-text to plain text, hOCR, TEI or eBook (PDF, MOBI, ePub).

Keywords: transcription, annotation, IIIF, OCR

Criteria for Recording and Categorizing Scholarly Digital Editions

Simon Sendler *¹, Daniela Schulz *²,

Hanna-Lena Meiners¹, Bernhard Fisseni³

¹ Göttingen State and University Library – Germany

² Herzog August Bibliothek Wolfenbüttel – Germany

³ Leibniz-Institut für Deutsche Sprache Mannheim – Germany

CLARIAH-DE is a project funded by the German Ministry of Research and Education, merging two original infrastructure projects (CLARIN-D and DARIAH-DE) concerned with the digital investigation of textual and linguistic sources within the humanities and cultural sciences. An important goal of CLARIAH-DE is to evaluate which representations are adequate for different kinds of data and data processing. The native format of an edition must be able to represent all relevant aspects of the edition. However, as different fields and methods of data processing employ different data formats, possibilities of transforming scholarly digital editions into various digital representations must be explored. This paper describes a procedure to provide initial recommendations for considering transformation potential based on information on the edition goals and data collected through a detailed online questionnaire.

Upon completing the questionnaire, respondents receive initial recommendations on data conversion into pivot formats and on the conversion of already existing digital editions to different native formats. Users are also offered contact information for further counseling and feedback as well as literature and web references. The recommendations of the survey provide a starting point and will need to be reflected on and adapted according to individual requirements.

Before developing recommendations, the authors determined appropriate criteria for assessing the complexity of a digital edition according to seven categories based on an informal evaluation of editions curated in the authors' home institutions, and in the DARIAH and TextGrid communities. The questionnaire is based on these criteria, and is designed for users without much experience with the various data formats available to scholars. Users can consult this questionnaire to examine the convertibility of their edition into a simpler pivot format to facilitate data exchange.

The questionnaire comprises seven questions ranging from the overall aim of the edition to type and number of source documents and desired features and depth of data capturing. Questions are close-ended and multiple choice. They accommodate different scholarly objectives. Each answer is assigned a numerical score related to the scholarly objective and the use(fulness) of different formats. The total score will offer an indication of the relative complexity of a digital edition, where complexity refers to both the variety and depth of various modes of annotation.

*Speaker

Questionnaire responses will be collected to document the use of the questionnaire and to assess the different needs of the community; collected data will also be used to improve the questionnaire. For this purpose, the open source software LimeSurvey hosted by the GWDG is being used.

The criteria as well as the mapping of criteria to recommendations in the questionnaire are reviewed, commented on, and documented by scholars involved in various current and recent scholarly digital editions, as well as scholars experienced in data conversion and annotation formats. Moreover, completed as well as ongoing edition projects will be explored and thus will help to validate the initial recommendations. Prospectively, the authors hope to extend the functionality of the questionnaire to also provide recommendations during the early stages of planning an edition.

Keywords:Digital Humanities, digital scholarly editions, interchangeability, standardization, interoperability, reuse, infrastructures

The CITE architecture – assigning knowledge to locations in the text

Neven Jovanovic *¹, Christopher Blackwell^{† 2}

¹ University of Zagreb, Faculty of Humanities and Social Sciences – Croatia

² Furman University – United States

Philology deals with texts. It attempts to fix their verbal expression, to understand their language, to interpret their contents. A common process underlies these activities: the process of making connections. When we review textual witnesses to select particular readings, we connect particular lexical forms with places in the text. When we comment on how we understand the language of a passage, we connect specific meanings to places in the text. And when we interpret a passage of a text, we connect certain significances to what has been said in places in the text. Our knowledge becomes material (explicit) so that it can be connected with equally "materialized" – clearly delimited and identified – parts of the text.

Digital philology does the same, only more so. Digital reading demands unambiguous identification; in exchange, it becomes (or can become) reproducible and falsifiable. In fact, our description of "philology" would be very different if it did not come from an attempt to model in a digital medium what a philologist does.

For such philology, we need to identify three things: places in the text, our comments, and connections between the two. This is never simple or straightforward. A "place in a text" can be a paragraph, a poetic line, a phrase, a syllable, a character. These places can overlap. They can be co-terminous without being identical ("of Achilles son of Peleus" might be read as one name, or two, as a half a hexameter, as a phrase in the genitive, as one adjective). Even the contents of a text change according to the particular reading (human or digital). Take this line of a play:

Creon: In season, all things are good. Oedipus: Do you know by what terms I will go?

Are "Creon" and "Oedipus" part of the text? For a human reader, these speaker-attributions are helpful. For an analysis of "the language of Sophocles" we probably do not need them. For an analysis of "the language of the characters in Sophocles" we might prefer something disjointed like "Creon: season", "Creon: good". Such transformations pose a challenge for citation.

A conceptual framework for a three-part system of textual references (location, connection, comment) is provided by the CITE architecture. It has been developed and implemented – with different software, different database models and different query language types – for more than a decade now by classical philologists Christopher Blackwell and Neel Smith. The CITE architecture defines schemes and protocols to a) address a segment of text in the most widely

*Speaker

[†]Corresponding author: cwblackwell@gmail.com

understandable manner; b) to preserve information about a position of the segment in the hierarchical structure of a text; c) to achieve this when a segment belongs simultaneously to several, perhaps even overlapping, structures.

We will present the current state of the CITE architecture, explain the URN identifiers at its core, demonstrate some of its applications, and comment on how CITE architecture is different from DOI, the Handle System and similar persistent identifiers.

Keywords: citation, referring, philology, URN, persistent identifiers

Approaches to Collaborative Multilingual Thesauri Development in Educational Context

Goran Zlodi ^{*†} ¹, Tomislav Ivanjko

¹ University of Zagreb, Faculty of Humanities and Social Sciences – Croatia

The paper discusses approaches, outcomes and experiences of an ongoing project of collaborative Croatian terminology contribution to the Art and Architecture Thesaurus (AAT) through the cooperation of university teachers, master level students and museum professionals. The project is conducted on the University of Zagreb Faculty of Humanities and Social Sciences (FHSS). A model proposed in the project can accelerate the otherwise time-consuming process of developing multilingual thesauri through greater student engagement while achieving multiple educational goals on real-world tasks. The methods of quality control and student engagement from crowdsourcing projects methodologies are applied, but they were revised and improved with respect to the specific needs and requirements of the educational context.

The process is segmented into the following stages:

1. Generating a corpus of the relevant terms.

Two thousand most frequent terms from four top facet (materials, techniques, objects and periods) were selected from the databases of partner heritage institutions.

2. Translation and linking.

Translation of terms into English and linking them to relevant AAT concepts (accompanied with references to relevant reference literature and lexicographical sources) is performed by students. Mapping of Croatian and English terms for specific concepts is further implemented through semantic technologies.

3. Quality control.

Methods include peer-checking among students and teacher supervision. Concepts that are particularly demanding (e.g. complex art techniques) are examined by a separate group of students, where examination also includes detailed literature research and comparison. The final stage of quality control is provided by scholars and museum professionals from the corresponding field.

4. Including the terms into AAT.

After the quality check, the terms would be included in the AAT and made openly available as Linked Open Data (LOD), provided by the Getty Research Institute. Further open formats for thesauri exchange (Zthes, RDF, JSON) would be provided. This will allow the reuse of project results by other national and international vocabulary projects and create a basis for enriching metadata and enabling multilingualism. That would increase the visibility, accessibility

*Speaker

†Corresponding author:gzlodi@ffzg.hr

and interoperability of diverse national heritage in an international context, thereby supporting multicultural communication at human and machine level.

During the mapping and validation of different concepts from monolingual and multilingual thesauri, via different methods and tools (SPARQL queries, LOD matching, using APIs), participants are becoming aware of linguistic and cultural differences in meaning and vocabulary construction (classifications, selection of preferred terms, fluid spatio-temporal boundaries etc.).

In this paper, we would like to discuss how participants (students, heritage professionals, scholars) perceive different aspects of vocabulary development and control and how they practice scholarly primitives (especially annotating, comparing, referring and representing) in the context of the ongoing developments in the methods and technologies related to controlled vocabularies and ontologies in the field of Digital Humanities.

Keywords: vocabularies, thesauri, crowdsourcing, education, scholarly primitives, linked open data, digital humanities

Modelling the socio-linguistic processes using the NLP tools for syntactic parsing, Neo4j Graph database for storing and semantic and communication analysis

Benedikt Perak * ¹

¹ University of Rijeka – Croatia

The paper deals with the application of the graph technologies for analysis of communication phenomena, primarily texts, using NLP tools, graph database and network algorithms. This approach produces tagged corpora stored in the custom model of the knowledge graph as labeled entities and property relations that can be used to explore and analyze various semantic relations within texts as well as their enriched extratextual information and correlations. More importantly for developing a humanistic type of research, this approach can be used to produce ontological model of the informational structures that can be queried across various custom based type types of entities connected to the linguistic references. This means that any linguistic level of morphosyntactic parsing can easily be related to various empirical analysis of the communication, conceptualization and framing of the social identities, interactions, institutions and cultural models.

As a case study, the paper will show the application of the graph technology for the analysis of the Croatian Parliament debates, covering sessions from the year 2003–2017. The data gathering process is published as a github project (<https://github.com/rodik/Sabor>). The debates of the 5th to 9th Parliamentary Assembly are downloaded as datasets of two types. In the original form, datasets are extensible but not connected in an information ontology that allows for an immediate extensive research with complex queries about the social, communicational or textual propensities. The data modelling and NLP parsing process is published in a paper by Perak and Rodik (Perak and Rodik 2018). The corpus contains data from 5 parliament Conventions, with 5599 discussion points, 895 members of the parliament, 42 political parties, 390 078 discussions. The goal of the research is to enable an extensive knowledge base with clearly structured ontology that will enable data integration, data enrichment, textual analysis and elaboration of quantitative-qualitative queries with multiple filters. We will present the process of the ontology creation using Python scripts and Neo4j database as well morphosyntactic parsing of the text using UDPipe NLP parser (Straka and Strakova 2017).

The rich social-conceptual structure of the discourse promoted by the speakers and political organizations is analyzed in terms of promoting certain perspective or a frame, using the influential words and analyzing their salience in terms of the conceptual network measures. In order to perform a Social Network Analysis (SNA) framework of the parliamentary and governmental texts, Python Igraph computational algorithms that explore the centralities and detect

*Speaker

communities of the lexical and communicational patterns (Perak and Ban Kirigin, in review) are used to identify, measure and visualize the topic analysis and conceptual framing promoted by various politically engaged speakers. For instance, using the coordinating construction it is possible to extract the most important concepts in the conceptual matrix of the lexeme mir "peace" and find the 5 (or more) speakers that most frequently used the concept, as presented in the illustration 1.

The application of the multidimensional graph approach can benefit empirical interdisciplinary type of research in the social sciences and humanities.

Keywords: parliamentary debates, NLP processing, corpus analysis, data integration, social, communication modelling

Mediating the Past

The Past is a Foreign Country: the Digital Rediscovery of the 19-c. Documents from the Metropolitan See of Philippopolis

Dimitar Iliev * ¹

¹ University of Sofia – Bulgaria

The paper presents the current work on the digital edition of the church chronicles of the Orthodox Metropolitan See of Philippopolis, or Plovdiv, in today's Bulgaria. This peculiar set of volumes, known as *kondika* in Bulgarian or *codix/codices* in Greek, is written in Greek and contains the annals of the region of Upper Thrace from the end of the 18th century until the late 1800s. This is a historical period of great turmoil, several wars and revolutions, the crisis of the Ottoman Empire and the creation of the modern nation-states on the Balkans. During the decades encompassed in the 4 volumes of these *codices*, the Episcopate of Philippopolis/Plovdiv passed from the Ecumenical Patriarchate in Constantinople to the newly-created Bulgarian Exarchate and fell within the boundaries of three successive states: the Ottoman Empire, the protectorate of Eastern Rumelia, and the Bulgarian Principality. In the chronicles of the See, we encounter information about taxes and donations, properties and changes thereof, marriages and deaths, the sustenance of monasteries, schools and hospitals in the region and their costs, local political, cultural and religious life and its protagonists. The digital edition of these annals, a contribution to DARIAH-EU on the part of the national Bulgarian consortium CLaDA-BG, suggests diverse activities from the editing of the original handwritten Greek text and its translation to the annotation of personal and place names, institutions and organizations, events and other entities as well as the creation of indices, vocabularies, notes and commentaries targeted at the scholarly community and the general public. The workflow is divided into several stages and includes researchers and assistants at various levels of their career. The present paper focuses on methodological challenges such as the recognition and annotation of different named entities in the text, the creation of indices that could serve as digital multilingual prosopographies and gazetteers of the historical period, and also the online publication of the annotated texts, original and translational, as a parallel aligned edition. The potential for further development of the digital edition in the context of the national consortium is also explained. Institutions such as the Ivan Vazov National Library of Plovdiv and the Ethnographical Museum to the Bulgarian Academy of Sciences are also content providers to CLaDA-BG. The former is currently conducting large-scale activities on the digitization of historical newspapers from the region of Plovdiv from the 1800s largely containing references to the same people, places and events that could be encompassed in a big searchable meta-collection together with the church annals. The latter is bound to provide materials to be included in CLaDA-BG's digital collection of historical artefacts from everyday life and popular culture which can also serve as illustrations to the notes in our digital edition about objects characteristic for 19-c. Thrace but virtually unknown to today's average user. Last but not least, the dataset extracted from our digital corpus will

*Speaker

be integrated into a detailed knowledge graph of our country and region from different periods derived from different sources.

Keywords: digital edition, parallel aligned texts, historical texts, prosopography, named entities, annotation

Integrating archival materials for the study of the turbulent Greek 40s

Vicky Dritsou^{*† 1,2}, Maria Ivanidou^{*1}, Isidora Despotidou², Vicky Liakopoulou², Karmen Vourvachaki², Panos Constantopoulos^{1,2}

¹ Digital Curation Unit, IMSI, Athena Research and Innovation Center (DCU) – Greece

² Department of Informatics, Athens University of Economics and Business (AUEB) – Greece

Humanities researchers often need to study heterogeneous digitized archives from different sources. But how can they deal with this heterogeneity, both in terms of structure and semantics? What are the digital tools they can use in order to integrate resources and study them as a whole? And what if they are unfamiliar with the methods and tools available? Towards this end, DARIAH-EU[1] and CLARIN[2] research infrastructures already support researchers in exploiting digital tools. Specific use case research scenarios have also been developed, with the PARTHENOS SSK[3] being a successful example. In this paper we describe our related (ongoing) experience from the development of the Greek research infrastructure APOLLONIS[4], where, among others, we have focused on identifying and supporting the workflows that researchers need to follow to perform specific research studies while jointly accessing disparate archives. Using the decade of 1940s as a use case, a turbulent period in Greek history due to its significant events (WWII, Occupation, Opposition, Liberation, Civil War), we have assembled (digitized) historical archives, coming from different providers and shedding light on different historical aspects of these events. From the acquisition of the resources to the desired outcome, we record the workflows of the whole research study, including the initial curation process of the digitized archives, the ingestion, the joint indexing of the data, the generation of semantic graph representations and, finally, their publication and searching. After the acquisition of the heterogeneous source materials we perform a detailed investigation of their structure and contents, in order to map the different archive metadata onto a common metadata schema, thus enabling joint indexing and establishing semantic relations among the contents of the archives. The next step is data cleaning, where messy records are cleaned and normalized. Natural Language Processing methods are then exploited for the extraction of additional information contained in the archival records or in free text metadata fields, such as persons, places, armed units, dates and topics, which enhance the initial datasets. The outcome is encoded in XML using the common schema and ingested into a repository through an aggregator implemented using the MoRE[5] system. A joint index based on a set of basic criteria is generated and maintained, thus ensuring joint access to all archival records regardless of their source. In addition, an RDF representation is generated from the encoded archival data, enabling their publication in the form of a semantic graph and supporting interesting complex queries. This is based on a specifically designed extension of CIDOC CRM[6] and a compilation of a list of research queries of varying complexity

*Speaker

[†]Corresponding author:v.dritsou@dcu.gr

encoded in SPARQL. Preliminary tests of the entire workflows and the tools used in all steps yielded very encouraging results. Our immediate plans include full scale ingestion and indexing of the material from a number of archives, producing the corresponding semantic graph and streamlining the incorporation of new archives.

DARIAH-EU, <https://www.dariah.eu/>

CLARIN, <https://www.clarin.eu/>

PARTHENOS Standardization Survival Kit (SSK), <http://www.parthenos-project.eu/portal/ssk-2>

APOLLONIS Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation, <https://apollonis-infrastructure.gr/>

MoRE Aggregator, <http://more.dcu.gr/>

CIDOC CRM, <http://www.cidoc-crm.org/>

e-Version of the Republic of Letters

Sepideh Alassi ^{*†} ¹, Lukas Rosenthaler ¹

¹ Digital Humanities Lab, University of Basel – Switzerland

Natural philosophers privately exchanged letters, manuscripts, and their newly printed books. The privacy of the letter allowed for the airing of unpopular and radically novel ideas, creating a mostly hidden discussion that carried on across Europe throughout the 17th century. This invisible *republic of letters* united like-minded thinkers and bridged the physical distances between them [1]. Correspondence analysis comprises a significant part of the history of science research as historians track the ideas of natural philosophers through their correspondence. Even if a researcher is interested in studying the contribution of one individual to a field, inevitably, she is led to examine the letter communications of others as well. That is because the correspondence data are by nature connected: natural philosophers either directly corresponded or mentioned each other's works.

As the tendency towards digitization increases, platforms emerge that individually present digital editions of correspondence categorized by authors. These platforms do not interoperate; as a result, the links between the data vanish, forcing the researchers using these platforms to study the textual sources entirely to find the relations [2]. Cataloging portals such as *Early Modern Letters Online* (EMLO) or the *Mapping The Republic Of Letters* help researchers with creating their research inventory by locating the resources on different platforms. However, these services only refer researchers to the sources of information without giving them direct access to the data. This paper presents an e-infrastructure that connects digital editions of the early modern scientific correspondence presented in different online platforms without re-creating data silos. Hence, through one single platform, researchers gain access to the semantically linked textual resources in a machine-readable format and can analyze the information using one set of research tools. Currently, the e-infrastructure brings three digital edition platforms to interoperate: the Bernoulli-Euler Online, The Newton Project, and the Briefportal Leibniz.

Representing the digital editions hosted on various endpoints as an RDF graph not only facilitates historians' research but also suggests a new research method - the network method. In contrast to the conventional approach that requires studying the individual resources to establish the connection between them, the network illustrates the links between resources, then having this picture in mind, researchers can explore the data. Furthermore, the e-infrastructure enables researchers to perform advanced searches on metadata of the network resources [2], or full-text searches for a phrase. This paper describes the asynchronous full-text search forwarding mechanism developed to this aim. Most of the digital edition platforms present text search functionality; the infrastructure benefits from this offered functionality enabling full-text searches on entire network resources without any need for collecting textual sources and indexing them. Furthermore, the paper describes an interactive web-based visualization tool that displays the RDF-based network of resources as 3D force-directed graphs [3]. This paper also presents a

^{*}Speaker

[†]Corresponding author: sepideh.alassi@unibas.ch

new feature recently added to introduce the time dimension to the visualizations. It allows researchers to study the change in the network over time, concerning both the topic and frequency of the correspondence.

Principe, Lawrence M., *The Scientific Revolution: A very short introduction*, Oxford University Press: Oxford, 2011.

Alassi, Sepideh et.al. *Newton virtually meets Euler and Bernoulli*, Digital Humanities 2019 Conference, Utrecht, Netherlands

Alassi, Sepideh et.al. *An Interactive 3D Visualization of RDF-based Digital Editions*, in review, Digital Humanities 2020 Conference, Ottawa, Canada

For information about the RDF standard, see <https://www.w3.org/RDF/>.

Keywords: e, infrastructure, RDF, interoperation, semantic digital editions, history of science, correspondence analysis, visualization

Infrastructural Challenges

CLARIAH-DE - Aligning two research infrastructures: Experiences and Challenges

Stefan Buddenbohm ^{*†} ¹, Dirk Goldhahn ²

¹ Göttingen State and University Library – Germany

² Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig – Germany

CLARIAH-DE is the merger of the research infrastructure networks CLARIN-D and DARIAH-DE. CLARIAH-DE is merging the two established research infrastructures over the course of two years (2019-2021). Thus, academic endeavours of researchers from the humanities and cultural sciences who work with complex digital tools and special data repositories shall be substantially facilitated. Both research infrastructures pursue goals directly connected to the scholarly primitives - discovering, annotating, comparing, referring, sampling, illustrating and representing - as they provide services and resources for them.

A large work package (WP4) is devoted to the harmonisation of the CLARIN-D and DARIAH-DE technical infrastructures. Infrastructure has to be seen both as basic or generic components like AAI, PIDs, monitoring, collaboration platforms but also as research oriented services, like the CLARIN Federated Content Search, the DARIAH Collection Registry or a GeoBrowser. Our work is based on previous cooperation, for instance the joint Technical Advisory Board (TAB), and preliminary integrative endeavours in recent years.

But despite well established communication processes such as the TAB it doesn't come as surprise that the practical merger leads to considerable challenges. CLARIN-D and DARIAH-DE have different disciplinary traditions, and have thus developed different technologies, tools, services and processes that now have to complement each other. The consolidation, however, is not straightforward, as is the case, in particular, with the merging of the three search and retrieval tools: Generic Search, Federated Content Search and Virtual Language Observatory. Consolidation often also requires the harmonisation of standards and interfaces (resource metadata, interchange formats). In other areas, such as Authentication and Authorization Infrastructure (AAI), we already found a solution based on CLARIN's and DARIAH's existing AAI-approaches. The work package is complemented by a technology watch that goes beyond CLARIN and DARIAH and tries to incorporate the most important developments in the field into its own plans wherever possible.

Beyond this we must also consider the European level. Both CLARIN-D and DARIAH-DE are contributing partners to the ERICs and limitations that have to be considered in the merger of two national infrastructures come with this interwovenness. CLARIAH-DE is directly involved in European development through its two partners. CLARIN-D is part of CLARIN - European Research Infrastructure for Language Resources and Technology and DARIAH-DE is the German partner of Digital Research Infrastructure for the Arts and Humanities (DARIAH-

*Speaker

†Corresponding author: stefan.buddenbohm@sub.uni-goettingen.de

EU).

In addition, and as research driven infrastructures, the community perspective is possibly most important, joining the picture of interdependencies: both CLARIN-D and DARIAH-DE are well established with users and dissemination activities which contributed to the level of trust researchers have in them. Trust is a very important and delicate asset which has to be taken into consideration at all stages of the merger process.

The paper presents some of our experiences in the merger process. We try to reflect not only on an infrastructural level but also involve the community perspective. We want to use this opportunity to collect feedback particularly from the European partners and projects and are interested in useful experiences from other initiatives.

Keywords: CLARIAH, DE, CLARIN, D, DARIAH, DE, merger, infrastructures, Germany

Services in a research infrastructure - an attempt to review current definitions and their implications

Laure Barbot ^{*†} ¹, Andrea Scharnhorst ^{* 1,2}

¹ Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

² Data Archiving and Networked Services – Netherlands

Services are one of the defining elements for research infrastructures, but how to define a service is less clear. In the broadest sense any support for research practices at different stages (as described in the Scholarly Primitives) can be perceived as a service. If we follow the definition of research infrastructures as knowledge infrastructures [1] then both support of the staff working in and with the infrastructure and all the equipment they work with potentially count as service. In DARIAH's strategic plan, services are defined exactly in this broad sense. As DARIAH pillars via which services are provided we find the Marketplace; Working groups; Foresight; and Education/Training [2]. In each of the pillars the provision of the service materialises by a combination of IT based components and other forms of support (such as enabling activities, knowledge exchange, network activities, funding etc.). What is further agreed upon is that (a) research infrastructures for post-modern research are usually distributed, networked structures and that (b) research infrastructures follow a complex architecture with different roles, stakeholders and perspectives involved [3]. For the daily practice of an ERIC as DARIAH is it important to shape a shared understanding how the high-level definition of the services (DARIAH pillar's) translates into concrete services as build in member countries (DARIAH contributions), by projects (pilots of services), in Working groups, and other actors inside of DARIAH (as by participating institutions). This paper explores different ways to define and order services, applying different dimensions (such as geographic spread, addressed knowledge domain, maturity, envisioned audience, etc.). We start from existing definitions of services as for the European Open Science Cloud, or IT services in general, zoom into current DARIAH-wide applied definitions (as from projects such as HaS, DESIR and SSHOC), and test them with existing services (as propagated via the DARIAH website, or collected in the DARIAH contribution tool). By a clearer understanding of the different nature of services we aim to support decision of what services to foster, innovate, and maintain on which level, by which means, and in which time frame.

Acknowledgement: This abstract bases on the draft White paper 'Towards a concise DARIAH service strategy' with contributions from DARIAH DCO and JRC members.

*Speaker

†Corresponding author: laure.barbot@dariah.eu

Edwards, P., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., ... Calvert, S. (2012). Knowledge Infrastructures: Intellectual Frameworks and Research Challenges. Report of a workshop sponsored by the National Science Foundation and the Sloan Foundation. <https://doi.org/http://hdl.handle.net/2027.42/97552>

Anonymous. (2019). DARIAH Digital Research Infrastructure for the Arts and Humanities - Strategic Plan 2019-2026. Berlin. https://www.dariah.eu/wp-content/uploads/2019/08/Strategic-Plan_2019-2026.pdf

Fihn, J., Gnad, T., Hoogerwerf, M., Jerlehag, B., Lenkiewicz, P., Priddy, M., & John Shepherdson. (2016). DASISH Reference Model for Social Sciences and Humanities Data Infrastructures (No. EC FP7 Grant Agreement Number: 283646). Brussels. (see <https://priddy.github.io/DARIAH-RA/introduction/rm-ssh/>)

Keywords: services, research infrastructure, DARIAH pillars, contributions

Digital libraries as part of the digital humanities infrastructure: services, tools and programmes

Renata Petrušić ^{*† 1}, Sofija Klarin Zadavec ^{*# 1}, Karolina Holub ^{*§ 1}

¹ National and University Library in Zagreb – Croatia

This paper will present the results of the analysis of services, tools and programmes that are offered by national libraries digital collections in order to support digital humanities research. Digital libraries are an important subject of interest for researchers in the digital humanities field, and their ability to be incorporated into digital humanities infrastructure is greatly influenced by professional, economic, legal, organizational and other factors. In doing so, digital libraries need to respond to the growing expectations and demands of researchers who are increasingly becoming participants in the development of the digital library systems. Collections of libraries are increasingly being viewed as collections as data and big data, indicating a shift from traditional libraries and the overall heritage sector to new forms and levels of work with the content of collections. National libraries are a particularly important part of digital humanities infrastructure because of their valuable collections, legally defined tasks and professional capabilities. The development of national digital libraries infrastructure involves the construction of a system for collecting digital legal deposit, harvesting web content, digitising resources, data aggregation, persistent identifiers, long-term preservation and various services. From the perspective of digital humanities researchers, the availability of digital resources with search and browse functions is only a minimum level of service, which should be followed by a range of functionalities tailored to each type of material, as well as the publication of datasets, the creation of data reuse tools (APIs), etc. The national libraries offer several forms of services and programmes for digital humanities researchers through data/digital laboratories, development of new tools through fellowship programmes and other methods of collaboration. The results of the analysis of the activities of national libraries in digital humanities research infrastructure will be compared to the activities of the National and University Library in Zagreb and further development of its services, tools and programmes will be presented.

Keywords: national libraries, digital libraries, digital humanities research, digital humanities infrastructure

*Speaker

†Corresponding author: rpetrusic@nsk.hr

#Corresponding author: sklarin@nsk.hr

§Corresponding author: kholub@nsk.hr

Synergy session

Geohumanities Data: Managing, Using, Opening, Sharing

James Smith *¹, Anna Barcz *², Erzsébet Tóth-Czifra *³,
Rebecca Kahn *⁴

¹ University College Cork – Ireland

² Ludwig Maximilian University [Munich] – Germany

³ Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

⁴ Alexander von Humboldt Institut für Internet und Gesellschaft – Germany

As the geohumanities or spatial humanities continue to grow and diversify, the tools of data ingestion, curation, management, visualisation and dissemination develop with them. The spatial turn is reaching a new maturity for humanists, and requires an increasing range of essential resources to undertake its core tasks. The study of space and place, and indeed the field of human and physical geography as a whole, is based on primitives of spatial knowledge: identity, location, magnitude, time as well as spatial entities such as the point, the line or the polygon. Through these spatial primitives, the languages of space and place emerge. These in turn inform a set of scholarly primitives: georeferencing, disambiguating, tabulating, managing, linking, archiving and so on. These interact with the corresponding primitives of data management: sustainability, interoperability, planning, legal compliance, standardisation, data sharing and so on. A new pipeline of spatial processes has emerged, with tools fitted to new primitives and accessible to an ever-widening range of humanists. The researcher can move through the processes of discovery and/or disambiguation (e.g. Wikidata or Pleiades), annotation (e.g. Recogito) and network analysis (e.g. Gephi or Palladio) or focus on one particular task, moving in and out of more traditional spatial systems such as GIS, itself now adapted to specific humanistic disciplines.

As Bodenhamer et al. (2010: vii) pointed out a decade ago, "humanists are fully conversant with space as concept or metaphor", but have been reviving "a dormant interest in the influence of physical or geographical space on human behaviour and cultural development". This trend, dormant no longer, has become the focal point of many digital humanities projects. Terms such as gazetteer, georeference, geoparser, shape file and GIS have become central to the vocabulary of many scholars. The goal of this synergy session between the Geohumanities and Research Data Management working groups is to consider the scholarly primitives of a new geohumanist norm: what are the data, how do humanists engage with them, how should they be managed,

*Speaker

and what tools, practices and collaborations - existing or incipient - best enable them to tell their story?

In a world of open scholarship, the FAIR principles, a plethora of data in a variety of formats and an increase in openly licensed and digitised content, getting back to basics - or primitives - is key. A certain level of disciplinary and infrastructural maturity calls for new norms and new workflows, as well as collaborations between the diverse working groups of DARIAH with a stake in the future of spatial research in the humanities.

Proposed format

Introduction (10 mins)

World cafe format thematic table discussions (20 mins)

Discussion based on summaries from rapporteurs (10 mins)

Audit of 'scholarly' primitives within Geohumanities and Research Data Management (10 mins)

Towards a data pipeline of Geohumanities tools and processes (10 mins)

Data sustainability in the Geohumanities (10 mins)

Discussion of synergies between WGs and proposals for future activities (20 mins)

****NB this session has been endorsed by the Geohumanities and Research Data Management working groups.****

Keywords:Data pipelines, Geohumanities, Interdisciplinarity, Research Data Management, Scholarly Primitives, Spatial Humanities, Spatial Turn, Sustainability, Synergy, Working Groups

DH process graphs

Panos Constantopoulos ^{*† 1,2}, Georgios Artopoulos ^{*3}, Costis Dallas ^{*1,4},
Piraye Hacigüzeller ^{*5,6}

¹ Digital Curation Unit, IMSI/Athena Research Centre – Greece

² Athens University of Economics and Business – Greece

³ The Cyprus Institute – Cyprus

⁴ Faculty of Information, University of Toronto (iSchool, UofT) – 140, St George street Toronto ON
M5S 3G6, Canada

⁵ University of Oxford – United Kingdom

⁶ Univeristy of Leuven – Belgium

In the light of Unsworth’s seminal work on scholarly primitives and early initiatives to understand the nature and capture the variety of digital methods and tools in the humanities through classifications (such as the AHDS Taxonomy of Computational Methods and the AHRC ICT Methods Network), the endeavour to understand and explicitly model scholarly processes found its place in the agenda of DARIAH as early as the preparatory phase. The Scholarly Research Activity Model was then formulated, followed by the NeDiMAH Methods Ontology (NeMO) which was developed within the NeDiMAH project and was subsequently adopted and supported by DARIAH VCC2 through the DiMPO WG. NeMO aimed at capturing the scholarly research process through a set of concepts representing the main elements of the humanities research ecosystem, their intrinsic structure, and the relations among them. The explicit representation of relations among concepts enables representations of research processes in the form of semantic networks best suited for associative, exploratory search and inference. Taxonomies, on the other hand, such as those above or TaDiRAH (developed by DARIAH-DE), can be incorporated as hierarchical term dictionaries in NeMO. The development of NeMO was informed by the extensive empirical study of scholarly information practices, needs and attitudes performed by the DiMPO WG across Europe, and was validated in a series of workshops. In subsequent work related to DARIAH-GR, a streamlined process (Research Spotlight) was developed to extract information from research articles, enrich it with relevant information from other Web sources, organize it according to the domain-neutral part of NeMO (appropriately elaborated, called Scholarly Ontology), and republish it in the form of linked data. This enables compiling semantic graph databases capturing who has done what, how, why and with what results. We

*Speaker

[†]Corresponding author: p.constantopoulos@dcu.gr

propose a synergy session in which the WGs involved will jointly explore the recording and analysis of actual and/or prototypical instances of work processes in their respective domains of interest and the compilation of semantic graph databases supporting advanced documentary and analytical work. The Digital Practices for the Study of Urban Heritage and the GeoHumanities WGs will put forward different classes of problems, yet all of them calling for heavy use of highly differentiated digital methods and corresponding goals. They will thus present complementary views on all aspects of the research process. For example, drawing on FRBRoo and the IfcOWL ontology, and going beyond previous efforts in linking Building Information Modelling elements with a conservation ontology, the proposed effort to bring together the WGs above will highlight the challenges of using a CIDOC CRM - based modelling approach to structure logically more diverse architectural data (i.e., structure, typology and usage) at a scale larger than a building, e.g., clusters of heritage buildings, forming streets and neighbourhoods of cultural and historical value. The Digital Methods and Practices Observatory WG will offer NeMO and its streamlined application as methodological framework. We expect the session to initiate collaboration among the WGs that will lead to an enriched conceptualization of their own domains, an appreciation of common (and non-common) patterns and, possibly, joint work on developing semantic graph compendium(s) of digital practices.

Keywords: research processes, digital methods, ontologies, digital practices, work analysis, semantic modelling

The value of collaboration. Challenges and opportunities for the Arts, Humanities and Social Sciences

Francesca Morselli ^{*1,2}, Andrea Scharnhorst ^{1,2}, Jennifer Edmond ,
Suzanne Dumouchel , Emiliano Degl'innocenti ³

¹ Data Archiving and Networked Services – Netherlands

² Digital Research Infrastructure for the Arts and Humanities – DARIAH-EU – France

³ Opera del Vocabolario Italiano (CNR-OVI) – Italy

Collaboration among different research domains and fields is paramount to innovation, and art, humanities and social sciences are no exception to the intensification of collaborative research practices (Wouters P., Beaulieu A., Scharnhorst A., Wyatt S., 2012). Science policy explicitly addresses the value of collaboration to reach innovative solutions to tackle the societal challenges as defined by Europe 2020 strategy or to meet the UNESCO Sustainable Development Goals (Pozzo, 2019).

In a relative small research area as the arts, humanities and social sciences (AHSS) this creates tension around the distribution of funds, and requires effective, conscious and agile ways to coordinate the action of Research Infrastructures (henceforth "RIs", or ERICs). Consequently, in the European Research Area, RIs follow this development, and collaboration is nowadays regarded more as a strategic requirement rather than something "nice to add" to a RI portfolio (ESFRI roadmaps 2018).

This synergy session will raise the awareness of existing collaborative mechanisms, as can be found at different levels of existing projects. More explicitly we will address:

The value of collaboration: what is the value and the purpose of collaborations among Research Infrastructures? Is it only to meet certain strategic demands, or do we really think this is a winning process, and if so for whom?

The future of collaboration: how will we face the future (ever more crowded) landscape together and apart? What are the future scenarios of collaboration between RIs? Which collaborative approaches should we incentivize?

The risks of collaboration: In this discussion the risks of collaboration - if any - will also be investigated. How can these be managed and leveraged on?

This workshop proposal aims to investigate these questions and to bring together representatives of research infrastructures in the fields of arts, humanities and social science to exchange their

*Speaker

views on the value of collaboration.

Rather than being an exchange on abstract principles of collaboration, we want the discussion to start around successful case studies of collaboration. Case studies can be large projects as well as small-scale endeavours, well-established activities or more experimental exchanges with different communities. The aim of this approach is twofold: on the one hand we wish to encourage the exchange and consequently the mutual learning from each others' experiences. On the other hand, starting from the empirical layer of the selected case studies, we wish to elicit what the strengths and weaknesses of these examples are. This approach will allow to draw general guidelines and recommendations to share with the AHSS community and beyond, and ultimately to develop ideas for new collaborations.

As a second innovative aspect this workshop will see the participation of RIs of different geographical dimensions: for example, regional and national infrastructures, in addition to the international infrastructures. The different granularity will encourage the discussion not only across disciplines, but also across geographic levels.

We are currently approaching representatives of RIs in AHSS to take part in this workshop, (such as OPERAS, E-RIHS, CLARIN) and we aim to have a variety of research infrastructures represented, both in terms of research fields (within the AHSS) and geographic area represented (international, national, regional).

Format

Round tables

Per table: 1 moderator and 1 person that takes notes

Schedule Session

Introduction presentation of selected case studies - 30 minutes

Round Tables discussion - 45 min

Report and Wrap up - 15 minutes

Keywords: research infrastructures, collaboration

