

Edited by Koraljka Kuzman Šlogar and Anamarija Žugić Borić

Digital Humanities & Heritage 1<sup>st</sup> and 2<sup>nd</sup> DARIAH-HR International Conference Zadar, Croatia – Rijeka, Croatia 13–15 October 2021 – 12–14 October 2022

#### Publisher

Institute of Ethnology and Folklore Research University of Zadar, Department of Information Sciences DARIAH-EU

#### For the publisher

lva Niemčić Josip Faričić Toma Tasovac

**Editors** Koraljka Kuzman Šlogar, Anamarija Žugić Borić

#### Proofreading Marta Brkljačić

**Layout** Vesna Beader

Cover design Anja Tkalec ISBN 978-953-8518-10-2

This book was funded with the support of the Ministry of Science, Education and Youth.

Zagreb – Zadar – Paris 2024

# **Digital Humanities & Heritage**

Proceedings of the 1<sup>st</sup> and 2<sup>nd</sup> DARIAH-HR International Conference Digital Humanities & Heritage Zadar – Rijeka, Croatia 2021 – 2022

> Edited by Koraljka Kuzman Šlogar and Anamarija Žugić Borić



Zagreb - Zadar - Paris 2024

### CONTENTS

To A N	<b>ma Tasovac</b> Note from the DARIAH Director	7
<b>Koraljka Kuzman Šlogar, Anamarija Žugić Borić</b> Introduction		9
1.	Susan Schreibman, Kelly Gillikin Schoueri 3D Digital Scholarly Editions: The Text as Object	15
2.	<b>Stefano Cursi, Letizia Martinelli, Filippo Calcerano,</b> <b>Michele Calvano, Luciano Cessari, Elena Gigliarelli</b> Digital Processes for the Conservation and Valorisation of Built Heritage	33
3.	<b>Mirna Willer</b> Relevance of "Scientific and Cultural Information" for Digital Humanities	59
4.	Daniel Ihrmark, Zoi Tsiviltidou Analysing Online Horror Fiction with Corpus Linguistic Techniques for Engaging Museum Labels: Creepypasta Wiki and the House of Terror's Mobile Application	73
5.	Natasa Thoudam The Gender of Al-Generated Arts in Sarah Kenderdine's "Deep Fakes" and Aishatu Gwadabe's Al Artivism	97
6.	<b>Danijela Birt Katić, Martina Krivić Lekić</b> Teachers as Creators of Ethnographic Knowledge – Presentation of Collected Material in a Digital Environment	115
7.	<b>Edward J. Gray</b> Adapting English-Language Pedagogy to Cultural Heritage and Digital Humanities in France: Results from using the IIIF Online Workshop with Future Research Engineers, Librarians, and Archivists	133
8.	Elena Lazarenko, Aleksandr Riaposov Community versus Corpus: Accessible Language Resources for Indigenous Speaker Communities as a Way to Preserve and Promote Cultural Heritage	145
9.	Laura Grzunov, Marta Ivanović, Marijana Tomić Enhancing Access to Digitised Glagolitic Material of Zadar County Through GlagoLab Portal	163

10. Martina Petrinović			
Participatory Mapping of the Cultural Heritage in the Mountain Region of Gorski Kotar in Croatia	185		
11. Miha Seručnik Slovenian Historical Topography: A Medievalist's Gazettee	r 199		

## A NOTE FROM THE DARIAH DIRECTOR

The application of digital methods in the traditional humanistic disciplines has ushered in a new era of scholarly exploration, preservation, and understanding. The digital humanities, a dynamic interdisciplinary field with a rich history of its own, dating back to the 1940s, has become pivotal in reshaping how we interpret and interact with our cultural heritage, and how we question and interpret our identities.

The DARIAH-HR International Conference Digital Humanities & Heritage stands as a testament to the diversity, creativity and innovative potential of humanistic inquiry using cutting-edge technologies. Organised by DARIAH-HR, the Croatian node of the pan-European Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU), this conference is becoming a focal point for researchers, students and practitioners not only from Croatia but also from Southeastern Europe and beyond.

DARIAH is a distributed research infrastructure, which means that, on the European level, we rely on the expertise and agility of our national nodes to build tools and services that can be used by scholars from across Europe and the world. And while DARIAH offers some central services – such as DARIAH-Campus<sup>1</sup> and SSH Open Marketplace,<sup>2</sup> to mention our two flagship platforms – it is national consortia, like DARIAH-HR, that bear the brunt of resource development, knowledge exchange and capacity building.

We live in the age of digital abundance, but the massive quantities of available digital content do not automatically translate into easy-touse research workflows. Digitally enabled, data-driven research in arts and humanities has the potential to influence public debates of great

<sup>&</sup>lt;sup>1</sup> https://campus.dariah.eu/.

<sup>&</sup>lt;sup>2</sup> https://www.sshopencloud.eu/ssh-open-marketplace.

social and cultural relevance, but arts and humanities researchers still face many challenges in finding, accessing, producing and reusing digital resources: filtering out the noise, processing metadata of different quality and granularity, clarifying reuse rights for digital objects, choosing the right standard to model data, identifying the best tools to process a myriad of data formats, etc.

This is why we need European research infrastructures and why we cannot possibly find solutions to all those challenges by working in isolation, within our national borders. Collaboration is key. Without public infrastructures such as DARIAH, we would be left at the mercy of commercial interests, who may not care as much as we do – as scholars and Europeans – about diversity, representation, accessibility, and sustainability.

Building infrastructures for arts and humanities is particularly challenging because arts and humanities are inherently interdisciplinary, involving different fields of knowledge and practice, as well as both qualitative and quantitative methods. Traditionally, the humanistic datasets were small, requiring close reading skills, and, while they still do, they are nowadays complemented by big data approaches, statistical analysis of digitised and born-digital materials, data visualisations and other kinds of so-called distant reading methods, all of which, together, help us develop a more comprehensive understanding of human culture, heritage, values, and identities.

I would like to take this opportunity to express my gratitude to Koraljka Kuzman Šlogar, the National Coordinator for DARIAH-HR, and her entire team for their tireless work on anchoring DARIAH in Croatia, as well as all the authors, reviewers, and participants who have contributed to making the DARIAH-HR International Conference *Digital Humanities* & *Heritage* a vibrant hub of intellectual exchange. May this volume inspire further dialogue, collaboration, and exploration. And may we all, sooner rather than later, meet again!

Toma Tasovac President of the DARIAH Board of Directors

### INTRODUCTION

#### KORALJKA KUZMAN ŠLOGAR, ANAMARIJA ŽUGIĆ BORIĆ

The field of digital humanities in Croatia is gradually evolving and gaining recognition, informally yet impactfully, as a dynamic area of scholarly inquiry, particularly following our nation's alignment with the founding members of the DARIAH-ERIC consortium. This alignment has catalysed a heightened awareness and propelled the advancement of digital humanities, fostering investment in educational initiatives, research endeavours, and the development of cutting-edge digital tools and methodologies across the humanities, social sciences, and arts. This burgeoning awareness underscores the transformative potential of digital humanities, enriching the academic community and illuminating its relevance to the broader public through its capacity to innovate and revolutionise traditional humanities scholarship.

However, a significant gap persists between researchers adhering to conventional methodologies and those embracing the evolving landscape of digital technologies, contemporary research methodologies, and advanced analytical and synthetic practices. Concurrently, there exists a notable divergence between humanists and cultural heritage professionals—archivists, museologists, and librarians—each group operating within its distinct paradigms. This plurality of perspectives, while presenting a challenge, also holds the promise of enhancing collective knowledge and fostering groundbreaking innovations in our scholarly pursuits.

The idea of initiating the Digital Humanities & Heritage (DHH) conference stemmed from a recognition of the imperative to establish a forum for interdisciplinary convergence, enabling scholars and experts from diverse backgrounds to engage collaboratively. Therefore, DHH is more than just an academic gathering; it serves as a nexus for collaborative discourse and intellectual exchange. By bringing together scientists, humanities scholars, and professionals from the fields of librarianship, archival science, and cultural resource management in museums, this conference illuminates the intricate interplay between digital humanities and cultural heritage. The DHH conference advocates for a paradigm shift, urging the adoption of digital technologies not merely as supplementary tools but as foundational methodologies essential for contemporary scholarship. Through embracing digital innovation, we equip ourselves to comprehensively understand, preserve, and explore our scientific, historical, and cultural heritage in unprecedented ways, adapting to the dynamic interplay of technology and human inquiry in our modern world.

The core principles guiding DARIAH's initiatives, as well as this scholarly conference, are firmly rooted in a commitment to the principles of open access to knowledge. This paradigmatic direction, which permeates all aspects of DARIAH's activities, represents a key component of our philosophy. Open access to knowledge underpins academic exchange and acts as a catalyst for fostering deeper understanding, innovation, and participation within the scholarly, artistic, and cultural communities. Consequently, this conference's primary mission extends beyond showcasing the latest advancements in digital humanities and heritage; it also seeks to elevate awareness and stimulate active engagement among all relevant stakeholders. By highlighting exemplary practices and sharing valuable insights, the conference aims to drive a shift toward an approach that is open, inclusive, and dynamic.

In this context, this scholarly gathering transcends being merely an academic event, evolving into a vibrant forum for inspiration, reflection, and collaboration. Through dialogue, presentations, workshops, and interactions, we aim to cultivate an environment that promotes the exchange of ideas and critically examines established paradigms. We recognize that open access to knowledge is not merely a theoretical ideal but an ethical imperative. In an era where disciplinary boundaries are increasingly fluid and the challenges we encounter grow more intricate, openness, transparency, and collaboration are essential tools for advancing progress.

Hence, the first edition of the DARIAH-HR international conference *Digital Humanities & Heritage* proceedings will be released as a digital edition, with a limited number of printed copies. It will be made available to all interested parties, including the academic community, and the general public through the DHH website and open access platforms. In line with our commitment to open access, this publication aims to promote the free exchange of knowledge and foster a deeper understanding of digital humanities when it is applied to cultural heritage domain. By making this work openly accessible, we ensure that cutting-edge research and innovative methodologies are available to a wider audience, fostering a more inclusive and collaborative academic landscape.

This volume includes selected presentations from the DHH2021 conference, held in Zadar, and the DHH2022 conference, held in Rijeka. It features an introductory note by Toma Tasovac, the President of the DARIAH-EU Board of Directors, and includes 11 papers that have undergone an international double-blind peer review process. These papers are appropriately categorised based on their scholarly discourse and relevance. The authors of these articles are scholars and experts from various parts of Europe and beyond. While all the works are interconnected by the theme of digital humanities, this collection has been organised according to more specific thematic focuses.

The first three articles thus explore the transformative impact of digital technologies on the preservation, interpretation, accessibility, and dissemination of cultural heritage. Each article addresses a distinct aspect of this broad theme, demonstrating how digital tools and methodologies are reshaping our understanding and management of cultural and scholarly information. The first article, "3D Digital Scholarly Editions: The Text as Object", authored by Susan Schreibman and Kelly Gillikin Schoueri, explores the innovative concept of treating 3D models as the central "text" in digital scholarly editions, emphasising their potential to create a new, multimodal space for research and annotation within the scholarly ecosystem. The second article, "Digital Processes for the Conservation and Valorisation of Built Heritage", authored by Stefano Cursi, Letizia Martinelli, Filippo Calcerano, Michele Calvano, Luciano Cessari, and Elena Gigliarelli, examines the impact of digital technologies on heritage sciences, particularly in the areas of survey, representation, modelling, and information management, highlighting the importance of preserving digital resources to prevent data loss and maintain the connection between past, present, and future. The third article, "Relevance of 'Scientific and Cultural Information' for Digital Humanities", authored by Mirna Willer, reflects on the significance of cultural heritage information within digital humanities, revisiting Ivo Maroević's concepts of scientific and cultural information and exploring their application through linked open data and Semantic Web technologies to enhance the organisation and dissemination of cultural heritage.

The next two articles investigate the intersection of digital narratives and artistic practices, highlighting how digital content and methodologies can enhance engagement and address broader societal issues within cultural contexts. The article "Analysing Online Horror Fiction with Corpus Linguistic Techniques for Engaging Museum Labels: Creepypasta Wiki and the House of Terror's Mobile Application", authored by Daniel Ihrmark and Zoi Tsiviltidou, explores how the language and content of a collection of top-rated born-digital horror narratives from Creepypasta Wiki can inform the writing of engaging and immersive museum labels for horror-related exhibitions, while the article "The Gender of AI-Generated Arts in Sarah Kenderdine's 'Deep Fakes' and Aishatu Gwadabe's AI Artivism", authored by Natasa Thoudam, compares the AI-generated art practices of Sarah Kenderdine and Aishatu Gwadabe, highlighting their distinct approaches and examining how their methods reflect the gendered relationship between women and technology.

The following two articles both investigate the preservation and dissemination of cultural heritage through educational initiatives, concentrating on how historical and ethnographic data are collected, processed, and made accessible using digital technologies. The article "Teachers as Creators of Ethnographic Knowledge – Presentation of Collected Material in a Digital Environment", authored by Danijela Birt Katić and Martina Krivić Lekić, examines the "Policies of Recording Ethnographic Materials from 1897 to 1954" (UČeka) project, focusing on archival research of manuscripts written by teachers to uncover the

key factors that influenced their ethnographic recording practices, and aims to popularise ethnology through a digital platform. The article "Adapting English-Language Pedagogy to Cultural Heritage and Digital Humanities in France: Results from using the IIIF Online Workshop with Future Research Engineers, Librarians, and Archivists", authored by Edward J. Gray, describes the Digital Technologies Applied to History (TNAH) Master's programme at L'École nationale des chartes, emphasizing the use of the IIIF Online Workshop to train students in the technical and English language skills necessary for preserving and valorising cultural heritage.

The following two articles showcase the potential of the digital environment in enriching and safeguarding endangered or extinct linguistic cultures. They emphasise the development of user-friendly digital platforms to facilitate the preservation, accessibility, and usability of valuable cultural and linguistic heritage, bridging the gap between expert communities and the general public. The article "Community versus Corpus: Accessible Language Resources for Indigenous Speaker Communities as a Way to Preserve and Promote Cultural Heritage", authored by Elena Lazarenko and Aleksandr Riaposov, accomplishes this task by directing its focus towards the adaptation of linguistic corpora for endangered Northern Eurasian languages into accessible web resources for the native speakers. On the other hand, the article "Enhancing Access to Digitised Glagolitic Material of Zadar County Through GlagoLab Portal", authored by Laura Grzunov, Marta Ivanović, and Marijana Tomić, focuses on improving the search and browsing interface of the GlagoLab Portal for digital collections of Glagolitic heritage of Zadar County, catering to both scholars and the general public.

The final two articles explain how participatory mapping and digital gazetteers can engage local communities and scholars in preserving cultural heritage, thereby enriching our understanding of both contemporary and historical landscapes. Martina Petrinović's article "Participatory Mapping of the Cultural Heritage in the Mountain Region of Gorski Kotar in Croatia" discusses how Croatian Society of Art Historians engages the local community in Gorski Kotar, Croatia, through a participatory mapping project to document and preserve cultural heritage and Miha Seručnik's article "Slovenian Historical Topography: A Medievalist's Gazetteer" explores the significance of historical gazetteers, illustrating the Slovenian Historical Topography database, which aids researchers in identifying and understanding place names and geographical features, especially in regions where multiple languages were historically spoken, such as present-day Slovenia.

Through Digital Humanities & Heritage, we invite you to join us in the exploration of the intersection between humanistic inquiry and technological advancement, where the past informs the present, and innovation paves the way for a richer, more inclusive understanding of our shared cultural heritage.

# **3D DIGITAL SCHOLARLY EDITIONS**: THE TEXT AS OBJECT

### SUSAN SCHREIBMAN

Maastricht University, Faculty of Arts and Social Sciences Maastricht, The Netherlands s.schreibman@maastrichtuniversity.nl https://orcid.org/0000-0002-8782-0794

### **KELLY GILLIKIN SCHOUERI**

Maastricht University, Faculty of Arts and Social Sciences Maastricht, The Netherlands k.gillikinschoueri@maastrichtuniversity.nl https://orcid.org/0000-0001-6350-2380

Digital scholarly editions have a long history in the area of digital scholarship. While the idea of treating a 3D model as the "text" of a digital scholarly edition is novel, it has the potential to re-imagine the role of these models within the scholarly ecosystem. Publishing 3D models as the central node within a scholarly knowledge site that includes apparatus, annotation, and paradata within an unbounded digital space, provides a new environment for 3D models to be studied within a multimodal and archontic space for inscription and investigation.

Keywords: digital Scholarly Editions, 3D modelling, digital scholarship, annotation, apparatus, paradata, long-term preservation, virtual research environment

#### INTRODUCTION

Digital scholarly editions were one of the first scholarly artefacts created by humanists after the development of the World Wide Web. The majority of these editions were text-based utilising the conventions of the Text Encoding Initiative (including The Rossetti Archive, The Walt Whitman Archive, The Thomas MacGreevy Archive, the Women's Writing Project, Documenting the American South), with a smaller number of image-based editions (such as The Blake Archive). The idea behind digital scholarly editions has been to use the technologies available to us to find new ways to present, analyse, and augment texts. Making texts available in digital format provides new modalities to contextualise, analyse and display, as well as, particularly over the past decade, allow the public the opportunity to collaborate in their creation. In short, a rationale for digital scholarly editions is to create a knowledge site for dissemination and research informed by the affordances of the medium (Pierazzo 2019).

The development of digital scholarly editions created an environment for a reconsideration of the purpose of text-based editions, a rethinking of a knowledge site around a text or a collection of texts within a constrained information space (e.g., the book) – in which annotation and contextualisation (apparatus) are delivered in an environment of compression and consolidation, to an unbounded digital space, potentially linking to the entirety of the World Wide Web (Gabler 2010). This shift allowed editors to no longer think about editions as sites of textual compression but as decentred webs of textuality. In Derridean terms, the digital archive is a space for gathering together texts on a specific theme, individual, or topic and expanding the world of its creation into a multimodal and archontic space for inscription and investigation (Schreibman and Papadopoulos 2019).

In the case of 3D models, a stable web-based environment, standards, and a common semantic language for the edition are lacking. Typically, these models are created by or for research teams for use in analytic and interpretive purposes. For example, in history, archaeology and architecture to better understand and communicate about the past (e.g., The War Poets Exhibition)<sup>1</sup> or test hypotheses of, for instance, building construction (e.g., Gruber and Dobbins 2013), light sources (e.g., Dawson et al. 2007), or visibility (e.g., Paliou et al. 2011). Once hypotheses or conclusions utilising the model are developed, researchers tend to write traditional scholarly articles in which the three-dimen-

<sup>&</sup>lt;sup>1</sup> http://ww1lit.nsms.ox.ac.uk/ww1lit/secondlife.

sional model is flattened into a two-dimensional plane and provided as an illustration. While 2D rendered images of 3D visualisations can be compelling, they can also be misleading, particularly in cases in which the realism and detail generated through realistic texture materials and accurate lighting simulations create an illusory nature as accurate representations of the past (Eiteljorg 2000; Clark 2010).

Some projects create a video of the 3D model or scene as a means of presenting their research in a more dynamic modality, often contextualised by an audio narration. These videos are frequently published on YouTube or Vimeo or displayed in museum exhibitions. While these derivatives are appealing due to their general ease of use and dissemination possibilities, the 3D models themselves do not realise their potential to become part of the (public) scholarly ecosystem (Schreibman and Papadopoulos 2019).

The presentation and preservation of web-based 3D models, augmented by annotation and apparatus, are still in their adolescence. This is opposed to text-based editions which have decades-old standards and research communities, such as the Text Encoding Initiative.<sup>2</sup> Currently, there are many 3D model formats, often developed for specific communities and sectors, limited software to display them, and browser support is variable (Champion 2016). While there have been several initiatives to address these issues (e.g., 3D-COFORM (Pitzalis, Kaminski, and Niccolucci 2011), CARARE (Hansen and Fernie 2010) and CS3DP (Moore, Rountry, and Scates Kettler (eds.) 2022)), there have been no agreed-upon community-based standards, particularly in terms of schema, metadata, and paradata standards. Thus the idea of developing a framework for digital scholarly editions in which the 3D model functions as the central "text" (be it a sculpture or an ancient artefact, a building, or a virtual world), to be annotated and contextualised within a single semantic and technical framework, is compelling. Building on traditions inherent in text-and-image-based editions, in which contextualising information is key to providing a more holistic experience of the text (in this case the model), including its relevance, providing information on its creation, as well as its historical, social, or cultural significance, is the objective of the

<sup>&</sup>lt;sup>2</sup> https://tei-c.org/.

PURE3D<sup>3</sup> project, begun in 2021 with a grant from The Dutch PDI-SSH (Platform for Digital Infrastructure for Social Sciences and Humanities).

# FROM SCHOLARLY EDITIONS TO 3D DIGITAL SCHOLARLY EDITIONS

The roles inherent in digital scholarly editions, particularly those that originate from analogue texts are fairly clear-cut. Taking as an example the novel Ulysses by Irish modernist writer James Joyce, it is clear that James Joyce is the author, and that Hans Walter Gabler is the editor of the edition of Ulysses published by Random House in 1984 (Joyce 1984). As editor of this edition, one of Gabler's roles was to put forth a theory of the text through the editorial process, correcting errors introduced during the authoring or publishing process, while providing a rationale for these edits, as well as augmenting the text with annotation to provide readers with context and background: historical, cultural, personal, and literary.

However, with editions of 3D models, if we think of the model as the text, who then is the author? If the model is a reproduction of, for example, a building, who is the author of the model of the building - in our analogy - the text? Is it the modeller who created the 3D surrogate? Or the architect who designed the analogue building? Or perhaps it is the builder who realised it? Or is it that none of these roles can be likened to the author of a novel? While we acknowledge that these parallels are not perfect, there is a case for thinking of the modeller (especially in the case of (re)constructions) as the author of the 3D text, especially when the modeller is recreating a building that is no longer extant (or is extant in parts or only at the foundation), or reconstructing an artefact from fragments, as there is much "authorial" intent (Brusaporci 2017, Gabler 2013, Tanselle 1976) that is brought to bear in the creation of the model. In instances of 3D modelling of objects or environments that no longer exist, or exist in some fragmentary form, the (re)constructions are much more speculative and can be used as heuristic tools from which to build knowledge about spatial and temporal relationships. For example, it is possible to use 3D models to simulate and analyse space in terms of light,

<sup>&</sup>lt;sup>3</sup> https://pure3d.eu/.

sound, or its usage by past human actors (Papadopoulos and Earl 2014) (Figure 1). These reconstructions can be compelling so that researchers look at evidence anew, or seek out evidence so that it can be built into the model and analysed within it (Wendrich et al. 2014). Thus, just as there can be many representations of a written text, edited according to different textual traditions, with editors interpreting the manuscript sources differently (Van Mierlo 2013), there can also be many interpretations of modelled physical artefacts created with different methods (e.g., photogrammetry, 3D laser scanning or structured light scanning), processed with different software (e.g., Agisoft's Metashape, RealityCapture or Meshroom) with modeller/editors making different choices during the digitisation and processing process according to theories of the text (e.g., the object's physical properties, such as scale, location, material, etc.). To further complicate the actors involved, given the possibility of more dynamic annotation (e.g., linked data and computer vision), the role of the editor may also be assumed by non-human actors.



Figure 1: A 3D (re)construction and luminance analysis of a potential lighting scenario in a prehistoric Greek ceramics workshop (© Papadopoulos and Earl 2014).

#### ANNOTATION FOR 3D DIGITAL SCHOLARLY EDITIONS

Digital scholarly editions in 3D share more with traditional text-based editions in the sense that they are assemblages or machines of knowledge as articulated by Deleuze and Guattari (2004). As such, they become open objects that can be read and understood through and by their means of production and reception. They can be used to revisit or re-elevate long-standing textual practices such as annotation, apparatus, and commentary as interlinked discourses, providing a palette for both textual

and contextual study. This model of thinking of editions repositions the editor, not as an objective arbiter of the text, but more like a hunter-gatherer, a constructor of a dynamic and rhizomatic knowledge site.

What can be represented in a 3D knowledge site? What constitutes an annotation? In 3D digital scholarly editions, in addition to text, the annotation may be multimodal and may contain linked data, reaching out to the web to bring into the edition knowledge created elsewhere. Additionally, one of the affordances of a dynamic 3D environment is the ability to position and change camera angles as a means of orchestrating and orientating viewing perspectives to call to attention (annotate) specific physical features of an artefact or its positioning spatially or temporally. Another annotative possibility is to layer multiple 3D objects within the scene, such as a map or other states of the object, configured in such a way that allows the user to enable and disable visibility. According to Huurdeman and Piccoli (2021), these types of implicit annotation (e.g., camera views, hypothesis layers, certainty index layers) generated via the user's interaction with the model and facilitated within the viewing environment, may be thought of differently from explicit annotation (e.g., text, image, sound, video) as they are less visible or recognisable as editorial interventions. As researchers, we have been schooled to create and understand the train of scholarship embedded in print-based annotative environments. These new modalities and environments for annotation raise new issues, such as how to make clear provenance, intention, and context, for both users of these editions, as well as for the scholarly record.

In the case of 3D (re)constructions, another form of implicit annotation can be the texture material used on the 3D model. For example, the type of floor that might have been in a Roman villa. Fragments of the floor remains from an archaeological excavation may indicate the general mosaic pattern but it is up to the modeller/editor to make an informed decision about the floor (re)construction based on known contemporary analogies to the site's region and period. Archaeologists have wrestled with these issues for many years, particularly as 3D technologies have improved in that we can create photorealistic models. In these cases, archaeologists have used different methods to signal the aspects of the representation which have been (re)constructed, such as the (re)constructed portion being digitally textured with a different material or colour as opposed to the known portions which have been discovered in-situ (Barratt 2021). This is in contrast to other schools of thought in which the idea is to make the (re)constructed portions of the structure blend seamlessly into the original, thereby creating a hyperreality in which real and hypothesised aspects are conflated in the service of perceived accuracy for a realistic experience (Gillings 2001). However, within a 3D digital scholarly edition, it is possible to present both representations. The model can be textured with 3D material that provides a representation in a pristine (re)constructed state and then reconfigured by the user with a click of a mouse to replace this pristine (re)construction with one that clearly shows which aspects of the building are original and which are (re)constructed. Moreover, the model could embed within it different (re)constructions - along with a confidence index to make clear how certain the modeller/editor is in her choices (Figure 2). This interactive content thus becomes part of the annotative structure. Similarly, audio annotation does not need to be limited to explicit annotation. Implicit audio annotation might be utilised for example, for a mediaeval church modelled to its original structure. The musicologist/modellers might speculate on the sonic spaces that would have existed millennia ago, to be triggered (dynamically or on-demand) based on location or point of view.



Figure 2: Certainty index colour scheme overlay of a (re)constructed Dutch Golden Age entrance hall, interactive through the Virtual Interiors Project prototype for a research-oriented 3D web viewer (© Huurdeman and Piccoli 2021).

In the case of archaeological research, archaeologists typically collect information during the excavation to document their process. Yet, in the act of uncovering remains, layers of evidence are destroyed (Barker 1986). Therefore, the excavation reports and notebooks are integral for future research because they contain the paradata, usually generated over years, documenting the excavation process. After an excavation campaign is completed, these notebooks are then often housed in museums or depots. These notebooks, reports and other documentation can also form part of the annotative structure of a 3D digital scholarly edition - either by selectively utilising images or text or via the digitisation of the entire notebooks and other evidence (photographs, drawings, videos, etc.) creating an assemblage, possibly stored (digitally and physically) in several locations - museums, archaeological depots, and the excavation site itself. Thus a goal of a 3D digital scholarly edition might be to create a virtual research environment allowing scholars and students who did not participate in the excavation to make sense of the vast amounts of materials collected. To this end, the role of the editor, as in text-based scholarship, is to select and organise so that the model can be read back through its creation, allowing other scholars to form their own hypothesis about the original site as well as its reconstruction.

The development and publication of digital editions need not be a linear process. The ability to easily revise or reconceptualise digital editions is one of its strengths. Huurdeman and Piccoli (2021) describe the development of 3D editions as a three-step iterative process. Rather than view the goal of the edition as a single output made available to the public at the end, it should be thought of as a staged process:

In our view, the use of 3D reconstructions as research tools enables a three-step (potentially iterative) process of knowledge creation, sharing, and discovery: (1) New insights are generated during the making of the 3D reconstruction (interpretative visualisation); (2) this new knowledge is presented to the viewer in an online platform (expressive visualisation); (3) the user gains, discovers, and creates additional knowledge through the interaction with the 3D environment and the exploration of the underlying (Linked) data.

Thus the goal is that the edition should supply users with the tools, apparatus and annotation to understand, contest, and learn from the representation (e.g., the model), as well as the decision-making processes of the (re)construction.

#### **DEVELOPING A 3D DIGITAL SCHOLARLY EDITION**

While the creation and online publication of a 3D scholarly edition can provide a rich, annotated and dynamically interactive 3D environment, there is still the technical challenge of holding all this information in one space. This is because a 3D digital scholarly edition is not simply, or only, what the user interacts with, it is also the infrastructure that supports it: the database and repository that houses the models, the annotation, and the metadata that enables access, interaction, and preservation of its constituent parts. Thus 3D digital scholarly editions can be conceived in terms of three interoperable tiers or layers:

- 1. The viewer layer (what the user sees);
- 2. The database layer that contains the version of the model used in the interface, as well as information contained in the annotation and information about the behaviour of these objects;
- 3. The preservation layer which can house all the versions of the model created, as well as the metadata and paradata.

As part of the PURE3D project, all three layers are included in its infrastructure. The viewing layer utilises software called Smithsonian Voyager<sup>4</sup> which is being developed by the Smithsonian Institution as a way to make publicly available the Smithsonian's collection of 3D digitised objects through online exploration and interaction in an immersive and engaging way<sup>5</sup> (Figure 3). The database layer is currently being developed by CLARIAH: the Common Lab Research Infrastructure for the Arts and Humanities<sup>6</sup> and will provide the access point for 3D scholarly editors to create, edit, and publish a 3D scholarly edition for their intended audience. Finally, the preservation layer is being informed by DANS: the Dutch Data and Archiving Networking Services<sup>7</sup> as an assurance that the 3D edition and associated data will be accessible long-term even if the viewing or database layers are no longer sustained.

<sup>&</sup>lt;sup>4</sup> https://smithsonian.github.io/dpo-voyager/introduction/.

<sup>&</sup>lt;sup>5</sup> View the Smithsonian 3D Collection at https://3d.si.edu/.

<sup>&</sup>lt;sup>6</sup> https://www.clariah.nl/.

<sup>7</sup> https://dans.knaw.nl/en/.



Figure 3: Voyager Explorer presentation of the Neil Armstrong Spacesuit (© Smithsonian).

While the life-cycle of a 3D edition is supported by the database and preservation layers, the development and conceptualisation of a 3D digital scholarly edition is principally concerned with the viewing layer – where explicit and implicit forms of annotation are integrated with a 3D model of historical or cultural significance. As a 3D web viewer, Voyager not only includes functionality for inspection of 3D assets (e.g., zoom in/out, measurement, cross-section, and visual rendering) but also focuses attention on system features that high-light aspects of digital storytelling through four distinct but interconnected modalities which provide for exploration both linearly and nonlinearly:

- 1. Annotation Labels: Short, expandable hotspot text as spatially aware annotation;
- 2. Articles: HTML-based pages with text and multimedia that can either overlay the 3D model or be situated to the side of it;
- 3. Guided Tours: A flexible combination of annotation, articles, camera movements and a set of analytic features, such as alternative material shaders, light settings measuring tape and slicer tool;
- 4. Audio Narration: A single audio clip that can be played at any time during the experience (Smithsonian 2023).

The Voyager Framework is appealing to the PURE3D project not only because of its rich feature capabilities both analytically and annotatively, but because the software is open-source, has robust documentation, and has a long-term commitment by the Smithsonian to update and maintain the code for web handling.

The first PURE3D project to be developed in Voyager was a building taken from the *Contested Memories*: The Battle of Mount Street Bridge.<sup>8</sup> This long-term research project takes as its inspiration a battle that ensued on and around Dublin, Ireland's Mount Street Bridge. This battle took place during the Irish 1916 Rising: a rebellion by the Irish against British colonial rule that took place in April 1916, beginning on Easter Monday, in the middle of the First World War. Although the British, with their superior resources, militarily and human, were able to put down the Rising within the week, it provided the impetus for Irish independence in December 1921.

The Battle of Mount Street Bridge was the first major encounter between a small band of Irish soldiers (17 in total) against two battalions of British soldiers who outnumbered them tenfold. What was significant about this battle, in a week of major defeats for the Irish, was that these 17 soldiers, who did not receive significant training, had a variety of mismatched guns and ammunition, and were not able to communicate with each other once they took their positions in four separate locations along and near Northumberland Road, were able to keep this much larger British force from crossing mount Street Bridge for most of the day. Eventually, the British forces prevailed, bringing in explosives and cannons, against which the Irish had little defence (Hughes et al. 2017).

The project began in 2013 with the principal research question of ascertaining the number of British casualties suffered during the battle. On the Irish side, this was clear – four Irish soldiers lost their lives. However, on the British side, the numbers varied widely from 150 to 350 casualties. The goal was to utilise the model to help provide a more definitive casualty figure. While an accurate number was determined: the British forces suffered 160 casualties (Hughes et al. 2017), this number

<sup>&</sup>lt;sup>8</sup> https://mountstreet1916.ie/.

was not arrived at via analytics within the model itself as we had originally presumed, but rather via an iterative research process as outlined above. Through this process of gathering sources and modelling them (in everything from spreadsheets to the virtual world itself), a different view of the battle was produced than what would have been generated if the intention of gathering sources was to write a traditional scholarly article. It was not until the research team engaged with the spatial affordances of the 3D scene that a different mental model of what transpired on that day in April 1916 emerged. We realised, however, that our beautiful model (at the time in the web-based Unity platform) did not allow those not involved in the project to gain the insights the project team did, or, indeed, arrive at alternative hypotheses. Thus, the idea of the 3D digital scholarly edition was born.<sup>9</sup>

The project team decided to explore the idea of a 3D digital scholarly edition using the first building the British troops encountered with Irish forces present, 25 Northumberland Road. As the edition was created with a general audience in mind, guided tours take a central role. The edition contains three tours: 1) the circumstances leading up to the battle; 2) information providing as full a picture as we can about the battle itself; and 3) the fall of 25 Northumberland Road and what happens to its combatants. Tours do not only provide a step-by-step narrative but provide a rich annotative environment of text and multimedia, including camera views, to help guide the reader's attention. For example, the narrative arc of the tours is underpinned by a map which locates the building in terms of other key locations (e.g., other buildings as well as the approach of the British) and facilitates a wider variety of camera positionings contextualised in space. This edition is intended to test and explore the limits of the Voyager system in terms of explicit and implicit annotative capabilities. The edition<sup>10</sup> benefits from a diverse range of multimedia resources collated from the previ-

<sup>&</sup>lt;sup>9</sup> While the Unity application offers a robust and flexible framework for developing 3D digital scholarly editions, there are some long-term preservation issues which preclude its suitability. For one, each unity project creates a bespoke interface and, therefore, a different semantic language on how to interact in the space is developed. Another issue is that a unity project may become obsolete or unusable when major web browsing updates occur. This would therefore require the expertise of a dedicated web developer to keep the Unity project viewable online.

<sup>&</sup>lt;sup>10</sup> https://pure3d.eu/25northumberlandrd/.

ous research from the "Contested Memories" project. These include short text summaries of the building, its combatants, and a timeline of events, as well as archival images, URL hyperlinks to further web-based sources, video and audio clips of first-hand accounts by the combatants, and finally 3D models of the two guns that were used to defend the building. These multimodal sources were developed as explicit annotations within Voyager's hotspot annotation and longer articles.

In terms of implicit annotation, the flexible capabilities of Voyager's guided tour allowed us to provide set camera views in and around the building. These camera movements, coupled with contextualising explicit annotation, provided the project team with a rich palette to indicate the various combatants' perspectives during the battle. For example, the two Irish Volunteers, Seamus Grace and Michael Malone were positioned at strategic locations within the building, one from a second-story room with three windows facing the front of the building and the other from a third-story bathroom window that was located on the side of the building (Figure 4). These locations provided Malone and Grace with optimal visibility and protection when the first company of British soldiers arrived at the crossroad juncture of Northumberland Road and Haddington Road (Figure 5).



Figure 4: Screenshot from 25 Northumberland Road edition displaying explicit annotations on the locations of the two Irish Volunteers positioned within the building during the conflict.



Figure 5: Left: Bird's eye view of the crossroad juncture of Northumberland Road and Haddington Road. Right: Vantage point of the two roads for Michael Malone from his third-story position (Tour 2, Step 5).

With both explicit and implicit annotation, the 3D model, underpinned by the map of the area, remained the central text upon which annotation was developed. We describe this as object-based annotation. In other words, annotation needs to have as its starting point a location on the 3D model. For example, the second tour, documenting the battle, opens with a view of the building from the perspective of the oncoming British soldiers. The next step of the tour reinforces this as it describes their journey from the port of Kingstown (now Dún Laoghaire) to the city centre. This annotation utilises documentary and archival sources to allow readers to understand the scope of the British forces brought in to put down the rising, as well as their reception along the journey. This annotation is augmented by a map that delineates the trajectory of their march for those not familiar with Dublin. In this way, throughout the edition, the 3D model is central to the narrative being told, providing both an anchor and a rationale for the annotative structure.

Our theory of the edition was tested during the testing phase of the project in Dublin in June 2022 (Schreibman and Gillikin Schoueri forthcoming). Overall, test study users felt that the inclusion of a variety of multimedia resources only added value to their contextual understanding of the battle, beyond simply providing text explanations. They also appreciated content which provided individual stories of the combatants as this humanised the encounter. As there exists a plethora of primary sources that provide first-hand accounts of the battle by combatants, we included quite a lot of this material. This was also appreciated as it enlivened the more neutral narrative provided by the editors. While test study users felt the explicit annotation was well-achieved through text and multimedia, they did, however, mention a desire for more use of the 3D space. For example, some users felt the set camera viewing perspectives could have shown more details of the surrounding environment or representations of the soldiers in the street, hence providing more implicit visual annotation to help users in their own interpretations of the environment. These comments reinforce the notion that implicit forms of annotation within a 3D digital scholarly edition can be strategically used to complement and enhance explicit editorial interventions. One comment that several individuals made during the feedback session was that they would have appreciated an audio narration of the text so they could listen to the annotation while exploring the 3D model. This will be developed and tested in future iterations of the edition.

#### CONCLUSION

A 3D digital scholarly edition is much more than the interface the user interacts with – it is the annotative apparatus surrounding that model and the insights gained as a result of the knowledge site. 3D Digital Scholarly Editions are in their infancy, but their promise is substantial in providing an environment for scholars to communicate the results of their research in a single conceptual space with the 3D model as the textual centre, and for users to interrogate and test their hypothesis within the virtual research environment created.

#### **REFERENCES LIST**

- Barker, Philip. 1986. Understanding Archaeological Excavation. London: B T Batsford Limited.
- Barratt, Robert P. 2021. "Speculating the Past: 3D Reconstruction in Archaeology". In Virtual Heritage: A Guide. Erik M. Champion, ed. London, UK: Ubiquity Press, 13–23. Available at: https://doi.org/10.5334/bck.c (accessed 12 October 2022).
- Brusaporci, Stefano. 2017. "The Importance of Being Honest: Issues of Transparency in Digital Visualization of Architectural Heritage". In Handbook of Research on Emerging Technologies for Architectural and Archaeological Heritage. Alfonso Ippolito, ed. Hershey, PA, USA: IGI Global, 66–93. Available at: https://doi.org/10.4018/978-1-5225-0675-1.ch003 (accessed 12 October 2022).

- Champion, Erik Malcolm. 2016. "The Missing Scholarship Behind Virtual Heritage Infrastructures". In Eurographics Workshop on Graphics and Cultural Heritage, GCH 2016. Chiara Eva Catalano and Livio De Luca, eds. Graz, Austria: Eurographics Association, 57–65. Available at: https://doi. org/10.2312/gch.20161383 (accessed 12 October 2022).
- Clark, Jeffrey T. 2010. "The Fallacy of Reconstruction". In *Cyber-archaeology*. Maurizio Forte, ed. Oxford, UK: Archaeopress, 63–73.
- Dawson, Peter, Richard Levy, Don Gardner, and Matthew Walls. 2007. "Simulating the Behaviour of Light inside Arctic Dwellings: Implications for Assessing the Role of Vision in Task Performance". World Archaeology 39 (1): 17–35. Available at: https://doi.org/10.1080/00438240601136397 (accessed 12 October 2022).
- Deleuze, Gilles and Félix Guattari. 2004. EPZ Thousand Plateaus. London, UK: A&C Black.
- Eiteljorg II, Harrison. 2000. "The Compelling Computer Image A Double-Edged Sword". Internet Archaeology 8. Available at: https://doi.org/10.11141/ia.8.3 (accessed 12 October 2022).
- Gabler, Hans Walter. 2010. "Theorizing the Digital Scholarly Edition". Literature Compass 7 (2): 43–56.
- Gabler, Hans Walter. 2013. "Editing Text-Editing Work". Ecdotica 10 (1): 42-50.
- Gillings, Mark. 2001. "Virtual Archaeologies and the Hyper-Real". In Virtual Reality in Geography. Peter Fisher and David Unwin, eds. London, UK: Taylor and Francis Group, 17–34.
- Gruber, Ethan and John Dobbins. 2013. "Illuminating Historical Architecture: The House of the Drinking Contest at Antioch". In Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, CAA2010. Francisco Contreras, Mercedes Farjas, and Francisco Javier Melero, eds. Oxford, UK: BAR Publishing, 71–76.
- Hansen, Henrik Jarl and Kate Fernie. 2010. "CARARE: Connecting Archaeology and Architecture in Europeana". In Digital Heritage. EuroMed 2010. Lecture Notes in Computer Science vol 6436. Marinos Ioannides, Dieter Fellner, Andreas Georgopoulos, and Diofantos G. Hadjimitsis, eds. Berlin, Heidelberg, Germany: Springer, 450–462. Available at: https://doi. org/10.1007/978-3-642-16873-4\_36 (accessed 12 October 2022).
- Hughes, Brian, Billy Campbell, and Susan Schreibman. 2017. "Contested Memories: Revisiting the Battle of Mount Street Bridge, 1916". British Journal for Military History 4 (1): 2–22. Available at: https://bjmh.gold. ac.uk/article/view/769 (accessed 12 October 2022).
- Huurdeman, Hugo and Chiara Piccoli. 2021. "3D Reconstructions as Research Hubs: Geospatial Interfaces for Real-Time Data Exploration of Seventeenth-Century Amsterdam Domestic Interiors". Open Archaeology 7 (1): 314–336.
- Joyce, James. 1984. Ulysses, corrected text edited by Hans Walter Gabler. New York, USA: Garland.
- Moore, Jennifer, Adam Rountry, and Hannah Scates Kettler, eds. 2022. 3D Data Creation to Curation: Community Standards for 3D Data Preservation. Chicago, IL, USA: Association of Research Libraries.

- Paliou, Eleftheria, David Wheatley, and Graeme Earl. 2011. "Three-Dimensional Visibility Analysis of Architectural Spaces: Iconography and Visibility of the Wall Paintings of Xeste 3 (Late Bronze Age Akrotiri)". Journal of Archaeological Science 38 (2): 375–386. Available at: https://doi.org/10.1016/j.jas.2010.09.016 (accessed 12 October 2022).
- Papadopoulos, Costas and Graeme Earl. 2014. "Formal Three-Dimensional Computational Analyses of Archaeological Spaces". In Spatial Analysis and Social Spaces: Interdisciplinary Approaches to the Interpretation of Prehistoric and Historic Built Environments. Eleftheria Paliou, Undine Lieberwirth, and Silvia Polla, eds. Berlin, Boston: De Gruyter, 135–166. Available at: https://doi.org/10.1515/9783110266436.135 (accessed 12 October 2022).
- Pierazzo, Elena. Digital Scholarly Editing: Theories, Models and Methods. Routledge, 2016.
- Pitzalis, Denis, Jamie Kaminski, and Franco Niccolucci. 2011. "3D-COFORM: Making 3D Documentation an Everyday Choice for the Cultural Heritage Sector". Virtual Archaeology Review 2 (4): 145–146. Available at: https://doi. org/10.4995/var.2011.4571 (accessed 12 October 2022).
- Schreibman, Susan and Costas Papadopoulos. 2019. "Textuality in 3D: Three-Dimensional (Re)Constructions as Digital Scholarly Editions". International Journal of Digital Humanities 1 (2): 221–233. Available at: https:// doi.org/10.1007/s42803-019-00024-6 (accessed 12 October 2022).
- Schreibman, Susan and Kelly Gillikin Schoueri. Forthcoming. "Voyages in 3D: Creating a Multimodal Narrative of the Battle of Mount Street Bridge". [Manuscript submitted for publication in the *Journal of Digital History*].
- Tanselle, G. Thomas. 1979. "The Editorial Problem of Final Authorial Intention". Studies in Bibliography 29: 167–211.
- Van Mierlo, Wim. 2013. "Reflections on Textual Editing in the Time of the History of the Book". The Journal of the European Society for Textual Scholarship 10: 133–161.
- Wendrich, Willeke, Bethany Simpson, and Eiman Elgewely. 2014. "Karanis in 3D: Recording, Monitoring, Recontextualizing, and the Representation of Knowledge and Conjecture". Near Eastern Archaeology 77 (3): 233–237.

## DIGITAL PROCESSES FOR THE CONSERVATION AND VALORISATION OF BUILT HERITAGE

### **STEFANO CURSI**

ISPC Institute of Heritage Science, National Research Council Montelibretti, Rome, Italy stefano.cursi@ispc.cnr.it

#### LETIZIA MARTINELLI

ISPC Institute of Heritage Science, National Research Council Montelibretti, Rome, Italy Ietizia.martinelli@ispc.cnr.it

#### FILIPPO CALCERANO

ISPC Institute of Heritage Science, National Research Council Montelibretti, Rome, Italy filippo.calcerano@cnr.it

#### MICHELE CALVANO

ISPC Institute of Heritage Science, National Research Council Montelibretti, Rome, Italy michele.calvano@ispc.cnr.it

#### LUCIANO CESSARI

ISPC Institute of Heritage Science, National Research Council Montelibretti, Rome, Italy Iuciano.cessari@ispc.cnr.it

#### **ELENA GIGLIARELLI**

ISPC Institute of Heritage Science, National Research Council Montelibretti, Rome, Italy elena.gigliarelli@cnr.it

The impact of digital technologies on heritage sciences has increased the speed and automation of processes and practices aimed at the survey, representation, modelling, and information management of the built, architectural, and archaeological heritage. These digital resources are progressively increasing, but only some of them are stored and disseminated, resulting in data loss, which means loss of connection between past, present and future and dispersion of memory and knowledge of the crafts, cultural traditions and *genius loci* related to the design and construction of historic buildings. This paper intends to illustrate, based on experiences in the field of built heritage modelling, the potential that a model-based approach can offer in order to build a main point of contact for architects, expert conservators, and researchers interested in implementing digital-based methodologies and strategies in the search for information models that can represent the uniqueness of the architectural heritage, and thus provide the ideal space in which data can take on a richer meaning and value, becoming an integral part of an information network that allows its correlation with different disciplines and fully reveals its value.

Keywords: built heritage, digital process, knowledge modelling, HBIM

#### INTRODUCTION

The built heritage is an expression of techniques and knowledge to be considered as a unique and irreplaceable source of aesthetic, historical, and cultural values, and therefore must be studied and documented in order to prevent any loss or damage, also ensuring that any restoration, maintenance, and reuse activities are undertaken consciously. Loss of data means loss of the link between past, present and future; loss of memory and loss of knowledge.

It should be noted that, in this field, we are dealing with a body of knowledge held together by experts who go on to investigate both the individual object and the artefact as an individual object, as well as its relationship with its context. Those activities are characterised by complex, multidisciplinary processes aimed at research, conservation and management of the architectural asset; during which every choice and action made by specialists is based on "what is known" about the asset and how that knowledge is represented and shared (Cursi, Simeone, and Toldo 2015).

Today, digital models play a key role in making historical buildings an integral part of contemporary urban life, fostering the convergence of past and present knowledge to facilitate their interpretation and preservation and supporting the prefiguration of their possible future use.

Digital transformation is having a significant impact on the information and communication technologies (ICT)-based methods and tools for the collection, analysis, production, interpretation, and dissemination of research and practices related to built heritage. This digital transformation has paved the way for important methodological changes, advancing documentation and conservation processes, favouring participative activities that bring together different experts with various technical or cultural backgrounds, and enhancing maintenance activities. Furthermore, the increasing ease of use of digital tools and the decreasing costs of instruments and software have favoured their diffusion to transdisciplinary communities of practice, increasing their application and experimentation in multiple fields and contexts.

A recent phase of this ancient journey intends to exploit digital technologies to create information systems capable of integrating the 3D representation of a historic building with evolving data produced during its life in a centralised, robust, coherent, and nonredundant repository, offering fundamental support for decision-making. These information systems can optimise and integrate conservation, restoration and enhancement processes, applied at different scales of intervention and for different purposes. However, despite the potential benefits, there is still much to be done in terms of methodological development, interoperability between software environments, open processes and tools, sharing of guidelines and best practices, and training and dissemination to the community of researchers and practitioners.

Based on the above, this paper aims, in the first instance, to offer a critical contribution to the cross-fertilising relationship between traditional approaches and the most advanced digitally-enabled research for the study, conservation, and valorisation of built heritage, as well as define an outline of the problems encountered in current approaches of digital information modelling for the built heritage, with particular attention to aspects concerning data exchange among different disciplines. Furthermore, it aims to illustrate the results of work completed and strategies pursued by work still in progress, conducted by the Built Heritage innovation Lab (BHiLab) at the Institute of Heritage Science (ISPC) of the National Research Council BHiLab research group, aimed at improving the semantic enrichment of digital models for the preservation and transmission of historical knowledge.

#### STATE OF THE ART

With the advent of digital transformation, survey activities have undergone a profound transformation, becoming a virtually unlimited container of information from which to extract different forms of representation (Herman et al. 2020).

Among the first experiences, we find the methodologies of 3D surveying and modelling and integration with 2D images aimed at creating a realistic digital model, called reality-based models (Cipriani, Bertacchi, and Bertacchi 2019; Jiménez Fernández-Palacios, Morabito, and Remondino 2017; Remondino (ed.) et al. 2018).

The main techniques to see considerable technological development are digital photogrammetry (Apollonio et al. 2021; Ulvi 2021) and 3D laser scanner acquisition (Buscemi et al. 2020; Hamal, Sari, and Ulvi 2020).

More recently, Historic Building Information Modeling (HBIM) (Murphy, McGovern, and Pavia 2009) has been recognised as a possible solution that can improve the representation of the built heritage and related knowledge (Logothetis, Delinasiou, and Stylianidis 2015; Pocobelli et al. 2018), mainly due to its ability to create a database containing a digital building description including geometry and other semantic information, such as components' classification and properties (Volk, Stengel, and Schultmann 2014).

At the architectural scale, it is possible to distinguish several strands of research on the use of HBIM for conservation and restoration. A pivotal theme is the stratigraphic analysis; a form of investigation done to reconstruct the history of a building, trying to understand what modifications were made and in what epochs. This, together with documents and written testimonies, allows us to reconstruct the will of the workers, the history of the building and the different phases of its life. To this end, Beltramo et al. (2019), building on the previous work conducted by Chiabrando et al. (2017) for mapping masonry decay in a BIM environment and by Diara et al. (2018) for the extraction and management of Industry Foundation Classes (IFC) data using Database Management System (DBMS), experimented, on the Abbey of S. Maria di Staffarda, an operational workflow for the integration of stratigraphic analysis in the HBIM environment.

For the representation of the Harris matrix, which is a methodology that uses diagrams to represent in an abstract way the temporal suc-
cession of soil layer formation, surface use, masonry construction, and their destruction, Mammoli et al. (2021) proposed a system for its semiautomatic construction and the mapping of stratigraphic units directly on the point cloud. More recently, Banfi et al. (2022) conducted work that aims to identify, in the BIM environment, stratigraphic surfaces and volumes capable of being mapped with descriptions and materials, allowing a bidirectional relationship between objects and information.

A second theme concerns the identification and semantic enrichment of architectural components. We can place within this strand the work conducted by Dore et al. (2015), which integrated building modelling created from historical data with procedural modelling developed from the laser scanner survey. Many researchers defined libraries of HBIM components using the resources made available by commercial software to create dedicated libraries of parametric objects (Garagnani, Gaucci, and Gruska 2016; Lopez et al. 2017; Oreni et al. 2014).

The new generation of tools that support the integration of data from laser scanners into BIM environments has led to the development of reality-based methodologies to overcome the representation's limitations of HBIM (Bolognesi and Garagnani 2018); although, to date, no algorithm has been developed to automatically convert point clouds into BIM components.

Other works focused on the creation and management of masonry abacuses in the HBIM environment. Specifically, Adami et al. (2020) proposed a new classification method, attempting to overcome the limitations of the traditional 2D system and exploiting the full potential of the solutions offered by BIM authoring tools.

Parallel to, but distinct from, the previous ones are the researches concerning abacuses of structural unrest. In this area, Zuccaro et al. (2010) proposed a multimedia self-training tool for the assessment of the damage caused by seismic events. We can also include the previously mentioned work of Dore et al. (2015) and that of Valero et al. (2018), who proposed a system for monitoring ashlar masonry walls of historic buildings by integrating reality capture tools, data processing (including machine learning) and HBIM models. The work of Bernardello et al. (2020), on the other hand, proposed two alternatives for creating

three-dimensional semantic objects in a BIM environment in which to integrate information usually embedded in two-dimensional drawings.

Furthermore, in the literature, there is a large body of research in the field of semantic data enrichment through the use of Semantic Web-related technologies (Pauwels, Zhang, and Lee 2017). Among these works, only a few have focused on the representation of architectural heritage. On the architectural scale, De Luca et al. (2011) have proposed the development of an information system that takes into account the relationships that can be defined between the representation of the building (shape, size, state of preservation, reconstructive hypothesis) and information, of a heterogeneous nature, from different fields (such as technical, documentary, or even historical), while a domain-specific ontology deals with the analysis of degradation, its diagnosis, and possible restoration interventions (Cacciotti, Blaško, and Valach 2015).

In addition, some research has focused on the integration of knowledge bases regarding architectural heritage, created through information ontologies, with a BIM environment (Acierno et al. 2017; Beetz et al. 2016; Maietti et al. 2018; Pauwels et al. 2013; Quattrini, Pierdicca, and Morbidoni 2017; Simeone, Cursi, and Acierno 2019). Recently, Yang et al. (2019) proposed a semi-automated mesh-to-BIM conversion through Dynamo and model translation to IFC owl, while Bonduel et al. (2020) investigated the applicability of Resource Description Framework (RDF) literals in a heritage context built to include a wide variety of existing geometric patterns. In addition, Borin et al. (2020) develop a model for visualising the Harris matrix (a methodology used for dating wall layers) through the use of a BIM-based ontology.

A further theme in the development of HBIM methodology deals with the representation and management of knowledge not directly traceable to building components. Several works demonstrate how these limitations can be overcome by integrating the HBIM environment with external resources and databases (Cursi et al. 2022).

Among the strategies for semantic enrichment of HBIM models, to extend model representation capabilities with dynamic data, we find experiences related to the Internet of Things (IoT) (La Russa and Santagati 2021; Masciotta et al. 2021; Shahinmoghadam and Motamedi 2019; Tang et al. 2019). The architecture envisioned for this line of research involves a physical deployment of sensors in the field, managed at the logical level by a layering of software capable of storing and making queryable the data produced by the sensors. This data will then be integrated into a database within which the HBIM model manages the positioning in space of the sensors and enables visualisation and querying.

## DIGITAL STRATEGIES FOR UNDERSTANDING, PRESERVING, AND SHARING CONSTRUCTIVE KNOWLEDGE

COLLABORATION AND KNOWLEDGE EXCHANGE IN BUILT HERITAGE PROCESSES

Based on the above, today, digital models play a very important role within the complex historical-archaeological framing of an artefact, because they foster convergence between knowledge of the past and the present and facilitate its interpretation.

From the experiences mentioned in the previous paragraph, it is evident how the past decade has witnessed a sharp increase in the use of digital models in the field of cultural heritage, fostered mainly by the technological evolution of digital surveying tools and the growing awareness of the potential of digital techniques in a field in which the public visibility and importance of cultural content determines its ideal testing ground.

There are currently many digital resources available on the market that are designed to improve design capabilities in specific areas and solve even very complex problems in well-defined subject areas (as in analysis, conservation and restoration projects, modelling, and simulation).

It is also true that the introduction of new technologies while helping to improve the processes of a field in a broad sense, brings with it a whole range of minor issues that contribute to increasing the complexity of the process. Although these software programs work well within the specific domain for which they were built, they do not help at all in making design collaborative, making it, on the contrary, more complex by increasing the distance between the communication and understanding of actors having different skills, while referring to the same objects on which they are working.

As already mentioned, however, the architectural heritage sector suffers from the scattering of this wide variety of digital data distributed in numerous independent archives, especially regarding unstructured artistic and cultural heritage. A further element of suffering is given by the great heterogeneity relating to the type of media and formats used for publication at accessibility levels and logical representation of models.

The greater number of disciplines involved and their intricate linkages lack coordination: the interdependence of activities, divergent design goals, and the difficulty of making known the reasons for one's design choices are just some of the problems that make life difficult for the designer (Borkowski et al. 2001; Garner and Mann 2003).

These causes are largely due to difficulties in communication and understanding of different models of reality that are used by actors to highlight aspects of their interests (Woo, Lee, and Sasada 2001). It is clear, however, that it is not only a problem of the quantity and correctness of the information exchanged but, more importantly, of its interpretation and the meanings associated with it (Gabriel and Maher 2002; Kalay 2001).

The exchange of reliable and consistent information among the actors in the process is always a necessary prerequisite for the successful outcome of the process (Carrara et al. 2015).

The lack of mutual understanding between such specialised tools is mainly due to issues related to the interoperability and semantic level of the ICT tools used, and the synchronisation methods they are forced to use to exchange information at a high level.

In fact, the proper formalisation of exchanged information and technical knowledge in general still remains an open problem: the excess of low-level information exchange, cause and effect of the potential of the new ICT, implies the simplification of the exchanged information, thus leading to misunderstanding among the actors and a step backwards in their effective communication. It is therefore evident how much the methodologies used to digitise and share the knowledge of an architectural asset turn out to be key nodes for it to be preserved and passed on.

As mentioned before, in recent times, the emergence of Building Information Modeling in the field of new construction has been matched by the interest of the scientific community in the search for new avenues to follow in defining a similar methodology applicable in the field of built heritage, now known by the acronym HBIM. The interest in this methodology is mainly due to its ability to create a repository of information, containing a digital description of the building that includes both 3D geometry and other semantic information, such as classification and properties of components (Volk, Stengel, and Schultmann 2014). This approach is based on the use of conceptual models typical of databases, with particular reference to the Entity-Relationship model. These models consist of first identifying some basic entities (such as the technological components of the building) and then identifying the relationships between these entities. The main computer applications are based, in this context, on relational databases that collect information according to well-defined standard schemas. A process, the latter, that allows the creation of a coherent conceptual model in which the central role is assumed by the represented object.

It is precisely this object-centred view, which is concerned only with describing its characteristics through a set of properties, that does not allow for the representation of semantically more complex associations and thus for the provision of a high enough level of formalisation to enable machines to perform tasks of understanding and inferring new information or linking different resources. In addition, complex metadata declarations related to provenance, context, reliability, and uncertainty of data are difficult or impossible to include in the model, despite being ubiquitous in heritage processes (De Luca et al. 2011; Gómez-Romero et al. 2015). Because BIM applications were originally developed for the design of new buildings, they often lack effective modelling tools to create geometries adherent to the needs of specialists involved in a process regarding the historic built environment. Moreover, in the case of new design, the need for accurate and complete knowledge of what we are going to model is crucial. This requirement aligns with the

information consistency imposed by BIM environments, which provide valuable support for the construction and management of a building. In contrast, considering the processes that characterise the activities of investigation and restitution of an existing asset, our knowledge on the asset often comes from extremely heterogenous sources. It is subject to continuous changes, interpretations, uncertanties, and gaps that must persist throughout the process and beyond. This limitation also extends to the relationships that are established between the components used, which, to simplify the work done by users, tend to impose constraints that reflect design principles and construction rules that are generally always valid for new construction but are also severe limitations in the documentation of historic buildings.

Therefore, works that led to the definition of innovative digital methodologies and technologies for the knowledge and documentation of historic buildings are illustrated, with the strategic objective of experimenting with a path that rediscovers, highlights, and traces the set of design and construction criteria of those artefacts, analysing the building techniques of the period and their specific application on the construction site, toward an approach projected into the entire building process.

In light of this, it then becomes essential to prefigure a modelling methodology in which both geometric and nongeometric aspects characterising a historic building can converge. As a consequence of the limitations imposed by the above systems, one possible solution is to integrate specific architectural heritage Knowledge Bases with the BIM environments.

Below, new modelling approaches and prototypes are illustrated that are the result of completed and ongoing research projects involving the authors of this paper being part of the BHiLab, together with other research institutions, universities, and industrial partners.

#### DIGITAL APPROACHES FOR BUILT HERITAGE MODELLING

Those illustrated below are some recent projects that tackle the abovementioned issues from different perspectives: the information system for the management and preservation of built heritage and cultural sites; the study of built heritage from an energy and environmental point of view; the integration of static and dynamic diagnostic and monitoring data within a 3D model; the structural analysis of unreinforced masonry structures; and experiments on the integration of ontological structures with HBIM models.

It is important to emphasise that knowledge of the building is a fundamental step and prerequisite in all the projects illustrated; an aspect that strongly influences both the way models are constructed and the results obtained depending on use.

The research conducted on the National Archaeological Museum of Naples (Museo Archeologico Nazionale di Napoli – MANN) (Martinelli, Calcerano, and Gigliarelli 2022; Calcerano et al. 2023), aims to support the planning and implementation of building documentation, maintenance, and conservation activities. To ensure that an HBIM model can effectively support this range of tasks, it should incorporate the results of analysis and expert knowledge that enable critical understanding of the building's construction system, its methods, materials, integration, and assembly of technological elements, and the different periods of construction of each part of the structure, to be represented within the HBIM model. In this way, the model assumes a central role in the building management and maintenance processes as a digital environment shared among specialists and specifically the managing authorities.

In this case, as mentioned earlier, data collection assumes a fundamental role in a solid understanding of a historic building, especially for an HBIM-based process aimed at representing the building's construction systems.

The analyses of the historical cities of Frigento (Italy) and Chuandixia (China), or Palazzo Maffei-Borghese (Rome, Italy) are instead characterised by the integration of the discipline of environmental design, through a simulation-based approach. The use of these tools as a nondestructive analysis technique, coupled with environmental monitoring and field surveys, allows deepening the comprehension of the environmental behaviour of built heritage, shedding new light on the understanding of ancient settlement principles, construction techniques, and design solutions (Gigliarelli et al. 2022; Gigliarelli, Calcerano, and Cessari 2016), a wealth of knowledge of the past to be preserved and shared. The research conducted as part of the project IDEHA – "Innovation for Data Elaboration in Heritage Areas", with the case of the Royal Site of Carditello, an 18<sup>th</sup>-century Bourbon palace near the city of Capua, presents a workflow solution to integrate environmental sensors and HBIM into an online platform, with the aim of developing a real-time data collection and monitoring services.

The results obtained, both from the visual analysis and from specific tests to investigate aspects of construction and studies of similar and/or coeval buildings, useful for identifying masonry stratigraphy or characterising deteriorated portions of the masonry, have been compiled within the model, which therefore assumes the role of common data environment (CDE).



Figure 1: Perspective section view from the inner courtyard of the HBIM model of the southwest wing of the National Archaeological Museum of Naples (project: HBIM4MANN, author: BHiLab-ISPC).



Figure 2: Royal Site of Carditello (IT). Methodology of modelling families through a first phase of visible analysis of the wall face and consultation of literature on coeval construction techniques in the Neapolitan area and a second phase of realisation of system families with the definition of the wall stratigraphy from a metric and textural point of view (project: IDEHA, author: BHiLab-ISPC).

An ad-hoc developed web interface will allow users who are not experts in BIM modelling to both view the model as well as access, query and manage the information collected and linked to it; both static information present in the model, and dynamic information coming from the sensors installed in the case study and traced back to the virtual counterparts. The solution is being tested in the case study with the monitoring of room temperature and humidity, in support of preventive conservation of the artworks exhibited there. This system will promote the sustainable enhancement of the built heritage through the use of innovative technological tools that will enable both the provision of integrated ICT systems to make heritage management, protection, restoration, and conservation activities sustainable, and the integration of cultural offerings with multi-channel front-end applications and services, including through mobile devices, that can generate added value for tourism, enhancement, and development of cultural sites.

With the HBIM4lazioHERITAGE project, BHiLAB is exploring the possibilities of integration between HBIM environments and IoT systems. The aim is to develop a prototype of a platform capable of integrating into real-time the data stream produced by sensors for user monitoring with the information model of the building housing the museum facility. The experimentation is taking place at the Bramante Cloister of the Church of Santa Maria della Pace, a building currently hosting the DART Museum in Rome, where exhibitions, temporary displays, and cultural events of contemporary art are organised.

In addition to the analysis of museum, visitor flows, again, the model will allow the storage, consultation and sharing of data resulting from the historical-architectural analysis that involved on-site study and in-desk research of bibliographic and archival records, historical maps and cartographies, and studies of similar and/or coeval buildings, providing a historical-critical guideline for other analyses.

Send-and-return data flow between the HBIM and IoT model is being tested in both of the above projects.

As part of the SISMI project – "Technologies for Safety Improvement and Reconstruction of Historic Centers in Seismic Areas", the work of Calvano et al. (2022) tests on a building in the historic centre of Cornillo Nuovo (Italy), a digital workflow aimed at connecting an HBIM modelling environment with the Quality Index of Masonry (IQM) method; developed in 2000 by the University of Perugia and updated by the Network of University Laboratories of Earthquake Engineering (ReLUIS). The proposed process involves overcoming the limitations of some rigidities due to BIM tools by using visual programming (VPL) for the analysis of the IQM.

Also in this experience, the construction of the system presupposed both direct and indirect surveys conducted on the building; information that was collected within the model. Specifically, particular attention is paid to the retrieval and organisation of data aimed at understanding the building's construction systems, such as the shape and dimensions of technical elements (especially when not visible from a geometric survey, e.g., the beam system of a wooden floor slab), construction nodes, etc.



Figure 3: Part of the code (Calvano et al. 2022) used to calculate the IQM expressed for the three stresses for the selected walls (project: SISMI. author: BHiLab-ISPC).

The work conducted on the monastery of St. Mary in Goranxi (Gigliarelli, Cangi, and Cessari 2022) focuses on the integration of a BIM environment with an artefact-specific knowledge base that aims to provide a solution that can fully cover the knowledge processed in a heritage process and necessary for its documentation and informed decisions regarding its preservation. At this stage, the proposed integration has been applied to the representation of some key heritage knowledge domains, namely typological components, structural components, temporal discontinuity components, masonry abacus, and structural failure abacus. The formalisation of knowledge related to these disciplines was integrated with a component-based representation provided by the BIM environment and necessary to control the geometric and technological characteristics of the architectural heritage artefact. In this process of integrating a BIM model with an ontological structure, a delicate aspect to be carefully considered is related to defining the right location in which to formalise information, as well as what information can be used to link the BIM area with the knowledge base to ensure interoperability between the two systems and to limit data duplication. As shown in Figure 4, the dialogue between these two environments is ensured by the transposition, in the ontology-based model, of the set of instances that make up the building model, guaranteed by a unique identifier that ensures one-to-one correspondence and that can be used as a reference for the development of ad-hoc solutions to transfer data from one environment to the other.



### • ONTOLOGY

Figure 4: The adoption of ontologies enables the enrichment of BIM semantics through a consistent, flexible, homogeneous and computable formalisation of both direct and indirect knowledge (project: "Multidisciplinary Technologies for the Study and Conservation of Post-Byzantine Monasteries in Southern Albania" (Tecnologie multidisciplinari per lo studio e la conservazione dei Monasteri Post Bizantini in Albania meridionale), author: BHiLab-ISPC).

The first advantage of this approach is related to the possibility of assigning different labels to a node representing a building component, ensuring its proper definition without being limited by the families defined in the BIM environment, usually constrained by the authoring software representation hierarchy. At the same time, technologies related to the Semantic Web allow multiple labels to be assigned to a node/object, and this is particularly useful in processes concerning the built heritage where building elements often take on different roles, and, consequently, refer to different classes, in the evolution of the building over time. All the data collected and interpreted by the experts were compiled into the ontological framework that allows the development of the building over time to be traced and interpreted from different points of view.

These works testify to the shift from a document-based information system, in which information about an artefact is stored in unstructured incremental repositories, to a model-based system, where each piece of information is part of a coherent and integrated, up-to-date representation. This faces several issues of current information systems, such as inconsistency or data duplication.

#### CONCLUSIONS

The results and experiences of the projects described above have solidified the conviction that it is necessary to systematically survey and systematise a specific body of knowledge in the field of historic building techniques, the loss of which can be traced back to the waning of "crafts", of traditional "making". The unavoidable purpose appears to be to record the various expressions of local history, tradition and technical culture that people have been able to express over centuries of building activity.

Conveying traditional knowledge into a computable digital environment allows for completely new pipelines to be followed for obtaining, preserving, sharing, and passing on constructive knowledge.

The experiences illustrated show how a model-based approach can offer clear advantages in the field of built heritage modelling. These solutions can facilitate the search, dissemination and sharing of content and help the understanding of cultural assets and the logical connections between works and places. A further step forward is to build a point of contact for architects, expert preservationists and researchers interested in implementing digital-based methodologies and strategies in the search for information models that can represent and witness the multifaceted richness of the uniqueness of built heritage, and promote discussion on how these digital approaches can be employed to improve knowledge, conservation, and sustainable management of architectural and archaeological heritage in the arts and humanities.

In this perspective, a research infrastructure such as DARIAH would provide fertile ground for building a working group that can become a Pan-European Virtual Hub, aimed at offering services, resources, guidelines and practices for the representation and management of knowledge collected and produced by research groups and specialists working in the field of historical, architectural, and archaeological built heritage.

The adoption of a Pan-European Virtual Hub will make it possible to reconstruct a new picture of the construction techniques spread around the world, the permanences and mutations of common typologies with the contribution of skilled workers who moved through Europe over the centuries. This will enable interdisciplinary exchange and interoperability, as well as help promote a culture of data sharing related to the permanence and mutations of common architectural types built over the centuries in Europe and a deeper engagement of heritage in modern society.

#### **REFERENCES LIST**

- Acierno, Marta, Stefano Cursi, Davide Simeone, and Donatella Fiorani. 2017. "Architectural Heritage Knowledge Modelling: An Ontology-Based Framework for Conservation Process". Journal of Cultural Heritage 24 (3): 124–133. Available at: https://doi.org/10.1016/j.culher.2016.09.010 (accessed 12 December 2022).
- Bruno, Nazarena and Riccardo Roncella. 2019. "HBIM for Conservation: A New Proposal for Information Modeling". Remote Sensing 11 (15): 1751. Available at: https://doi.org/10.3390/rs11151751 (accessed 10 December 2022).
- Apollonio, Fabrizio Ivan, Filippo Fantini, Simone Garagnani, and Marco Gaiani. 2021. "A Photogrammetry-Based Workflow for the Accurate 3D Con-

struction and Visualization of Museums Assets". Remote Sensing 13 (3): 486. Available at: https://doi.org/10.3390/rs13030486 (accessed 7 April 2022).

- Banfi, Fabrizio, Raffaella Brumana, Angelo Giuseppe Landi, Mattia Previtali, Fabio Roncoroni, and Chiara Stanga. 2022. "Building Archaeology Informative Modelling Turned into 3D Volume Stratigraphy and Extended Reality Time-Lapse Communication". Virtual Archaeology Review, 13 (26): 1–21. Available at: https://doi.org/10.4995/var.2022.15313 (accessed 7 November 2022).
- Beetz, Jakob, Ina Blümel, Stefan Dietze, Besnik Fetahui, Ujwal Gadiraju, Martin Hecher, Thomas Krijnen, Michelle Lindar, Martin Tamke, Raoul Wessel, and Ran Yu. 2016. "Enrichment and Preservation of Architectural Knowledge". In 3D Research Challenges in Cultural Heritage II. How to Manage Data and Knowledge Related to Interpretative Digital 3D Reconstructions of Cultural Heritage. Sander Münster, Mieke Pfarr-Harfst, Piotr Kuroczyński, and Marinos Ioannides, eds. Cham: Springer Cham, 231–255. Available at: https://doi.org/10.1007/978-3-319-47647-6\_11 (accessed 7 April 2022).
- Beltramo, Silvia, Filippo Diara, and Fulvio Rinaudo. 2019. "Evaluation of and Integrative Approach Between HBIM and Architecture History". In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. GEORES 2019 2nd International Conference of Geomatics and Restoration. Raffaella Brumana, Valeria Pracchi, Fulvio Rinaudo, Alberto Grimoldi, Marco Scaioni, Marco Previtali, and Luigi Cantini, eds. XLII-2-W11: 225–229. Available at: https://doi.org/10.5194/isprs-archives-XLII-2-W11-225-2019 (accessed 7 April 2022).
- Bernardello, Rachelle Angela, Paolo Borin, Federico Panarotto, Andrea Giordano, and Maria Rosa Valluzzi. 2020. "BIM Representation and Classification of Masonry Pathologies Using Semi-Automatic Procedure". In Brick and Block Masonry – From Historical to Sustainable Masonry. Jan Kubica, Arkadiusz Kwiecień, and Łukasz Bednarz, eds. London: CRC Press, 771–778.
- Bolognesi, Cecilia and Simone Garagnani. 2018. "From a Point Cloud Survey to a Mass 3D Modelling: Renaissance HBIM in Poggio a Caiano". In IS-PRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2: 117–123. Copernicus GmbH. Available at: https://doi.org/10.5194/isprs-archives-XLII-2-117-2018 (accessed 17 October 2022).
- Bonduel, Mathias, Anna Wagner, Pieter Pauwels, Maarten Vergauwen, and Ralf Klein. 2020. "Including Widespread Geometry Schemas into Linked Data-Based BIM Applied to Built Heritage". Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction, 172 (1): 34–51. Available at: https://doi.org/10.1680/jsmic.19.00014 (accessed 17 May 2022).
- Borin, Paolo, Rachele Angela Bernardello, and Anna Grigoletto. 2020. "Connecting Historical Information with BIM Ontologies. HBIM Methods for the Visualization of Harris Matrix for the Torrione in Carpi". In *Graphi*-

cal Heritage. Volume 1 – History and Heritage. Luis Agustín-Hernández, Aurelio Vallespín Muniesa, and Angélica Fernández-Morales, eds. Cham: Springer, 757–770. Available at: https://doi.org/10.1007/978-3-030-47979-4\_65 (accessed 17 May 2022).

- Borkowski, Adam, Cherif Branki, Ewa Grabska, and Wojciech Palacz. 2001. "Towards Collaborative Creative Design". Automation in Construction 10 (5): 607–616. Available at: https://doi.org/10.1016/S0926-5805(00)00070-4 (accessed 17 May 2022).
- Buscemi, Francesca, Pietro Maria Militello, Cettina Santagati, Marianna Figuera, Graziana D'Agostino, and Damiano Antonino Angelo Aiello. 2020. "Use and Reuse of Spatial and Quantitative Data in Archaeology: From 3D Survey to Serious Game at Phaistos (Crete)". Archeologia e Calcolatori 31 (1): 189–212. Available at: https://www.archcalc.cnr.it/indice/ PDF31.1/09\_Buscemi\_et\_al.pdf (accessed 7 April 2023).
- Cacciotti, Riccardo, Miroslav Blaško, and Jaroslav Valach. 2015. "A Diagnostic Ontological Model for Damages to Historical Constructions". *Journal* of *Cultural Heritage* 16 (1): 40–48. Available at: https://doi.org/10.1016/j. culher.2014.02.002 (accessed 7 April 2023).
- Calcerano, Filippo, Stavroula Thravalou, Letizia Martinelli, Kristis Alexandrou, Georgios Artopoulos, and Elena Gigliarelli. 2023. "Energy and Environmental Improvement of Built Heritage: HBIM Simulation-based Approach Applied to Nine Mediterranean Case-studies". Building Research & Information 52 (1–2), 225–247. Available at: https://doi.org/10.1080/09 613218.2023.2204417 (accessed 7 April 2023).
- Calvano, Michele, Letizia Martinelli, Filippo Calcerano, and Elena Gigliarelli. 2022. "Parametric Processes for the Implementation of HBIM–Visual Programming Language for the Digitisation of the Index of Masonry Quality". ISPRS International Journal of Geo-Information 11 (2): 93. Available at: https://doi.org/10.3390/ijgi11020093 (accessed 7 April 2023).
- Carrara, Gianfranco, Antonio Fioravanti, Gianluigi Loffreda, and Armando Trento. 2015. Conoscere Collaborare Progettare: Teoria Tecniche e Applicazioni per la Collaborazione in Architettura. Rome: Gangemi Editore spa.
- Chiabrando, Filiberto, Massimiliano Lo Turco, and Fulvio Rinaudo. 2017. "Modeling the Decay in an HBIM Starting from 3D Point Clouds. A Followed Approach for Cultural Heritage Knowledge". In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2-W5: 605–612. Available at: https://doi.org/10.5194/isprsarchives-XLII-2-W5-605-2017 (accessed 7 May 2024).
- Cipriani, Luca, Silvia Bertacchi, and Gianna Bertacchi. 2019. "An Optimised Workflow for the Interactive Experience with Cultural Heritage through Reality-Based 3D Models: Cases Study in Archaeological and Urban Complexes". In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2-W11: 427-434. Available at: https://doi.org/10.5194/isprs-archives-XLII-2-W11-427-2019 (accessed 7 May 2024).
- Cursi, Stefano, Letizia Martinelli, Nicolò Paraciani, Filippo Calcerano, and Elena Gigliarelli. 2022. "Linking External Knowledge to Heritage BIM".

Automation in Construction 141 (September): 104444. Available at: https://doi.org/10.1016/j.autcon.2022.104444 (accessed 7 May 2024).

- Cursi, Stefano, Davide Simeone, and Ilaria Toldo. 2015. "A Semantic Web Approach for Built Heritage Representation". In *Computer-Aided Architectural Design Futures*. The Next City New Technologies and the Future of the Built Environment. Gabriela Celani, David Moreno Sperling, and Juarez Moara Santos Franco, eds. Part of the book series *Communications in Computer and Information Science* 527: 383–401. Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: https://doi.org/10.1007/978-3-662-47386-3\_21 (accessed 17 May 2022).
- De Luca, Livio, Chawee Busayarat, Chiara Stefani, Philippe Véron, and Michel Florenzano. 2011. "A Semantic-Based Platform for the Digital Analysis of Architectural Heritage". Computers & Graphics, Virtual Reality in Brazil, 35 (2): 227–241. Available at: https://doi.org/10.1016/j.cag.2010.11.009 (accessed 17 May 2022).
- Diara, Filippo and Fulvio Rinaudo. 2018. "Open Source HBIM for Cultural Heritage: A Project Proposal". In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2: 303–309. Copernicus GmbH. Available at: https://doi.org/10.5194/isprsarchives-XLII-2-303-2018 (accessed 17 May 2022).
- Dore, Conor, Maurice Murphy, Stephen McCarthy, F. Brechin, C. Casidy, and Evelien Dirix. 2015. "Structural Simulations and Conservation Analysis – Historic Building Information Model (HBIM)". Available at: https://doi. org/10.5194/ISPRSARCHIVES-XL-5-W4-351-2015 (accessed 4 April 2021).
- Gabriel, Gerard Cesar and Mary Lou Maher. 2002. "Coding and Modelling Communication in Architectural Collaborative Design". Automation in Construction 11 (2): 199–211. Available at: https://doi.org/10.1016/S0926-5805(00)00098-4 (accessed 4 April 2021).
- Garagnani, Simone, Andrea Gaucci, and Bojana Gruska. 2016. "Archaeo-BIM: An Innovative Method for Archaeological Analysis of an Etruscan Temple in Marzabotto". In Arqueológica 2.0. 8th International Congress on Archaeology, Computer Graphics, Cultural Heritage and Innovation. José Luis Lerma and Miriam Cabrelles, eds. Valencia: Editorial Universitat Politècnica de València, 314–317. Available at: https://cris.unibo.it/ handle/11585/563007#.XxRfuHtxdPY (accessed 17 May 2022).
- Garner, Steve and Phebe Mann. 2003. "Interdisciplinarity: Perceptions of the Value of Computer-Supported Collaborative Work in Design for the Built Environment". *Automation in Construction*, Computer Aided Architectural Design Research in Asia, 12 (5): 495–499. Available at: https://doi.org/10.1016/S0926-5805(03)00035-9 (accessed 17 May 2022).
- Gigliarelli, Elena, Filippo Calcerano, and Luciano Cessari. 2016. "Implementation Analysis and Design for Energy Efficient Intervention on Heritage Buildings". In Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. Marinos Ioannides, Eleanor Fink, Antonia Moropoulou, Monika Hagedorn-Saupe, Antonella Fresa, Gunnar Liestøl, Vlatka Rajčić, and Pierre Grussenmeyer, eds. Part of the book

series Lecture Notes in Computer Science 10058. Cham: Springer International Publishing, 91–103. Available at: https://doi.org/10.1007/978-3-319-48496-9 (accessed 17 May 2022).

- Gigliarelli, Elena, Giovanni Cangi, and Luciano Cessari. 2022. "Rilievo per la modellazione e la gestione informativa HBIM. Approccio multicomponente per l'analisi strutturale e il restauro di edifici storici". Archeologia e Calcolatori 33 (1): 135–155. Available at: https://doi.org/10.19282/ ac.33.1.2022.08 (accessed 7 May 2024).
- Gómez-Romero, Juan, Fernando Bobillo, Montserrat Ros, Miguel Molina-Solana, M. Dolores Ruiz, and Maria J. Martín-Bautista. 2015. "A Fuzzy Extension of the Semantic Building Information Model". *Automation in Construction* 57 (September): 202–212. Available at: https://doi.org/10.1016/j. autcon.2015.04.007 (accessed 7 May 2024).
- Hamal, Seda Nur Gamze, Binnaz Sari, and Ali Ulvi. 2020. "Using of Hybrid Data Acquisition Techniques for Cultural Heritage: A Case Study of Pompeiopolis". Türkiye İnsansız Hava Araçları Dergisi 2 (2): 55–60.
- Herman, Grigore Vasile, Tudor Caciora, Dorina Camelia Ilies, Alexandru Ilies, Anca Deac, Amalia Sturza, S. M. Sonko, Norbert-Szabolcs Suba, and S. Nistor. 2020. "3D Modeling of the Cultural Heritage: Between Opportunity and Necessity". *Journal of Applied Engineering Sciences* 10 (1): 27–30. Available at: https://doi.org/10.2478/jaes-2020-0005 (accessed 7 May 2024).
- Jiménez Fernández-Palacios, Belen, Daniele Morabito, and Fabio Remondino. 2017. "Access to Complex Reality-Based 3D Models Using Virtual Reality Solutions". Journal of Cultural Heritage 23 (January): 40–48. Available at: https://doi.org/10.1016/j.culher.2016.09.003 (accessed 7 May 2024).
- Kalay, Yehuda E. 2001. "Enhancing Multi-Disciplinary Collaboration through Semantically Rich Representation". Automation in Construction, Design Representation, 10 (6): 741–755. Available at: https://doi.org/10.1016/ S0926-5805(00)00091-1 (accessed 7 May 2024).
- La Russa, Federico Mario, and Cettina Santagati. 2021. "An AI-Based DSS for Preventive Conservation of Museum Collections in Historic Buildings". *Journal of Archaeological Science*: Reports 35 (February): 102735. Available at: https://doi.org/10.1016/j.jasrep.2020.102735 (accessed 7 May 2024).
- Logothetis, Sotiris, Anika Delinasiou, and Efstratios Stylianidis. 2015. "Building Information Modelling for Cultural Heritage: A Review". ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W3 (August): 177–183. Available at: https://doi.org/10.5194/isprsannals-II-5-W3-177-2015 (accessed 7 May 2024).
- Lopez, Facundo, Pedro Lerones, Jose Llamas, Jaime Gómez-García-Bermejo, and Eduardo Zalama. 2017. "A Framework for Using Point Cloud Data of Heritage Buildings Toward Geometry Modeling in A BIM Context: A Case Study on Santa Maria La Real De Mave Church". International Journal of Architectural Heritage 11 (May): 965–986. Available at: https://doi.org/10. 1080/15583058.2017.1325541 (accessed 4 April 2021).
- Maietti, Federica, Marco Medici, Federico Ferrari, Anna Elisabetta Ziri, and Peter Bonsma. 2018. "Digital Cultural Heritage: Semantic Enrichment and

Modelling in BIM Environment". In Digital Cultural Heritage. Marinos Ioannides, ed. Part of the book series Lecture Notes in Computer Science 10605. Cham: Springer International Publishing, 104–118. Available at: https://doi.org/10.1007/978-3-319-75826-8\_9 (accessed 4 April 2021).

- Borin, Paolo, Rachele Angela Bernardello, and Anna Grigoletto. 2020. "Connecting Historical Information with BIM Ontologies. HBIM Methods for the Visualization of Harris Matrix for the Torrione in Carpi". In Graphical Heritage. Volume 1 – History and Heritage. Luis Agustín-Hernández, Aurelio Vallespín Muniesa, and Angélica Fernández-Morales, eds. Cham: Springer, 757–770. Available at: https://doi.org/10.1007/978-3-030-47979-4\_65 (accessed 17 May 2022).
- Mammoli, Raissa, Chiara Mariotti, and Ramona Quattrini. 2021. "Modeling the Fourth Dimension of Architectural Heritage: Enabling Processes for a Sustainable Conservation". *Sustainability* 13 (9): 5173. Available at: https:// doi.org/10.3390/su13095173 (accessed 7 May 2024).
- Martinelli, Letizia, Filippo Calcerano, and Elena Gigliarelli. 2022. "Methodology for an HBIM Workflow Focused on the Representation of Construction Systems of Built Heritage". *Journal of Cultural Heritage* 55 (May): 277– 289. Available at: https://doi.org/10.1016/j.culher.2022.03.016 (accessed 4 April 2021).
- Masciotta, Maria Giovanna, Maria José Morais, Luís F. Ramos, Daniel V. Oliveira, Luis Javier Sánchez-Aparicio, and Diego González-Aguilera. 2021. "A Digital-Based Integrated Methodology for the Preventive Conservation of Cultural Heritage: The Experience of HeritageCare Project". International Journal of Architectural Heritage 15 (6): 844–863. Available at: https://doi. org/10.1080/15583058.2019.1668985 (accessed 4 April 2021).
- Murphy, Maurice, Eugene McGovern, and Sara Pavia. 2009. "Historic Building Information Modelling (HBIM)". *Structural Survey* 27 (4): 311–327. Available at: https://doi.org/10.1108/02630800910985108 (accessed 4 April 2021).
- Oreni, Daniela, Raffaella Brumana, Andreas Georgopoulos, and Branka Cuca. 2014. "HBIM Library Objects for Conservation and Management of Built Heritage". International Journal of Heritage in the Digital Era 3 (June): 321–334. Available at: https://doi.org/10.1260/2047-4970.3.2.321 (accessed 4 April 2021).
- Pauwels, Pieter, Danilo Di Mascio, Ronald Meyer, and Rens Bod. 2013. "Integrating Building Information Modelling and Semantic Web Technologies for the Management of Built Heritage Information". In 2013 Digital Heritage International Congress (DigitalHeritage). IEEE. Available at: https:// doi.org/10.1109/DigitalHeritage.2013.6743787 (accessed 4 April 2021).
- Pauwels, Pieter, Sijie Zhang, and Yong-Cheol Lee. 2017. "Semantic Web Technologies in AEC Industry: A Literature Overview". Automation in Construction 73 (January): 145–165. Available at: https://doi.org/10.1016/j. autcon.2016.10.003 (accessed 7 May 2024).
- Pocobelli, Danae Phaedra, Jan Boehm, Paul Bryan, James Still, and Josep Grau-Bové. 2018. "BIM for Heritage Science: A Review". Heritage Science 6 (1): 30. Available at: https://doi.org/10.1186/s40494-018-0191-4 (accessed 4 April 2021).

- Quattrini, Ramona, Roberto Pierdicca, and Christian Morbidoni. 2017. "Knowledge-Based Data Enrichment for HBIM: Exploring High-Quality Models Using the Semantic-Web". *Journal of Cultural Heritage* 28 (November): 129–139. Available at: https://doi.org/10.1016/j.culher.2017.05.004 (accessed 7 May 2024).
- Remondino, Fabio, Diego Gonzalez-Aguilera, Panagiotis Agrafiotis, and Andreas Georgopoulos, eds. 2018. Latest Developments in Reality-Based 3D Surveying and Modelling. Basel: MDPI Multidisciplinary Digital Publishing Institute. Available at: https://directory.doabooks.org/handle/20.500.12854/51448 (accessed 7 May 2024).
- Shahinmoghadam, Mehrzad and Ali Motamedi. 2019. "Review of BIM-Centered IoT Deployment: State of the Art, Opportunities, and Challenges". ISARC Proceedings (May): 1268–1275.
- Simeone, Davide, Stefano Cursi, and Marta Acierno. 2019. "BIM Semantic-Enrichment for Built Heritage Representation". Automation in Construction 97 (January): 122–137. Available at: https://doi.org/10.1016/j.autcon.2018.11.004 (accessed 7 May 2024).
- Sungho Woo, Eunjoo Lee, and Tsuyoshi Sasada. 2001. "The Multiuser Workspace as the Medium for Communication in Collaborative Design". Automation in Construction, CAADRIA, 10 (3): 303–308. Available at: https:// doi.org/10.1016/S0926-5805(00)00046-7 (accessed 7 May 2024).
- Tang, Shu, Dennis R. Shelden, Charles M. Eastman, Pardis Pishdad-Bozorgi, and Xinghua Gao. 2019. "A Review of Building Information Modeling (BIM) and the Internet of Things (IoT) Devices Integration: Present Status and Future Trends". Automation in Construction 101 (May): 127–139. Available at: https://doi.org/10.1016/j.autcon.2019.01.020 (accessed 7 May 2024).
- Ulvi, Ali. 2021. "Documentation, Three-Dimensional (3D) Modelling and Visualization of Cultural Heritage by Using Unmanned Aerial Vehicle (UAV) Photogrammetry and Terrestrial Laser Scanners". International Journal of Remote Sensing 42 (6): 1994–2021. Available at: https://doi.org/10.1080/01 431161.2020.1834164 (accessed 3 May 2024).
- Valero, Enrique, Alan Forster, Frédéric Bosché, Camille Renier, Ewan Hyslop, and Lyn Wilson. 2018. "High Level-of-Detail BIM and Machine Learning for Automated Masonry Wall Defect Surveying". Conference: 35th International Symposium on Automation and Robics in Construction (ISARC), Berlin (July): 740–747. Available at: https://doi.org/10.22260/ ISARC2018/0101 (accessed 7 May 2024).
- Volk, Rebekka, Julian Stengel, and Frank Schultmann. 2014. "Building Information Modeling (BIM) for Existing Buildings — Literature Review and Future Needs". Automation in Construction 38 (March): 109–127. Available at: https://doi.org/10.1016/j.autcon.2013.10.023 (accessed 7 May 2024).
- Yang, Xiucheng, Yi-Chou Lu, Arnadi Murtiyoso, Mathieu Koehl, and Pierre Grussenmeyer. 2019. "HBIM Modeling from the Surface Mesh and Its Extended Capability of Knowledge Representation". ISPRS International Journal of Geo-Information 8 (7): 301. Available at: https://doi. org/10.3390/ijgi8070301 (accessed 4 April 2021).

Zuccaro, Giulio, Francesco Cacace, and M. Rauci. 2010. "MEDEA: A Multimedia and Didactic Handbook for Structural Damage and Vulnerability Assessment – L'Aquila Case Study". In *Summary Report of the Cooperative* Activities of the COST Action C25, Sustainability of Constructions. L. Braganca, H. Koukkari, R. Landolfo, V. Ungureanu, E. Vesikari, and O. Hechler, eds. Malta: Department of Civil & Structural Engineering, Faculty for the Built Environment, University of Malta, 2 (2011): 201–213.

### RELEVANCE OF "SCIENTIFIC AND CULTURAL INFORMATION" FOR DIGITAL HUMANITIES

#### MIRNA WILLER

University of Zadar, Department of Information Sciences (retired professor) Zadar, Croatia willer.mirna@gmail.com

The paper offers some reflections on the annual seminars Archives, Libraries, Museums: Possibilities for Co-operation Within the Environment of the Global Information Infrastructure (AKM, Rovinj-Poreč, Istria 1997–2019, 2021), and explores the relevance of the cultural heritage information sector's place in the complex of digital humanities. The concepts of scientific and cultural information within the theoretical field of information sciences elaborated by Maroević and essential to the organisation of information within the cultural heritage sector are argued to be as relevant today as when they were presented at the 2<sup>nd</sup> AKM seminar in 1998. These concepts are analysed within the context of principles and models of linked open data and Semantic Web technologies on the example of the work in progress on the new IFLA's standard ISBDM: International Standard Bibliographic Description for Manifestation.

Keywords: archives, libraries, museums' seminars; cultural heritage; scientific information; cultural information; ISBDM; linked open data; semantic web; digital humanities

#### INTRODUCTION

The paper opens with an overview of the history of and underlying ideas behind cooperation between heritage institutions in the context of annual seminars Archives, Libraries, Museums: Possibilities for Co-operation Within the Environment of the Global Information *Infrastructure* (1997–). This context covers areas of theoretical issues primarily within the field of information sciences, cutting-edge technology and state-of-the-art in standards development, issues of information organisation in the design of finding aids and their application in information systems and user services, and presentation of current practices and projects within a given topic.

The central part of the paper focuses on discussions about the need to revisit the relevance of information-finding aids in times of technological change. As part of the theoretical discussion of the topic at the 2<sup>nd</sup> seminar in 1998, Professor Ivo Maroević called for a rethinking of the concepts of "scientific and cultural information". The validity of Maroević's ideas is further confirmed within the context of principles and models of linked open data and Semantic Web technologies on the example of the work in progress on the new IFLA's standard ISBDM: International Standard Bibliographic Description for Manifestation in the 2020s.

The conclusion wraps up a notion of continuity of rethinking the theoretical background behind principles and models that govern the design of standards and information-finding aids in the digital humanities sector and recognising their interdependence with the current information technologies.

#### ON SEMINARS ARCHIVES, LIBRARIES, MUSEUMS

The annual seminars Archives, Libraries, Museums: Possibilities for Co-operation Within the Environment of the Global Information Infrastructure (Arhivi, knjižnice, muzeji: AKM, Rovinj-Poreč, Croatia 1997–2019, 2021), have been organised around the idea that libraries, archives and museums are faced with the same, common challenges in order to rethink and transform their functions and services within the new paradigm imposed by the all-pervading information and communication technologies. The complexity of the problem had made it quite clear that an interdisciplinary approach in which experts from these three communities would learn from and teach each other was not enough, and that active cooperation with the ICT sector, academia, and sectors like publishing, cultural tourism and, indeed, cultural industries was needed. To accomplish that, it was necessary to rethink the theoretical framework of information sciences that would accommodate the impact of new technologies in the lifecycle of an information cultural object and to analyse, test, and harness that technology in the services libraries, archives, and museums provide for their present and prospective users.

The organisers of the annual seminars are three national associations - Croatian Archive Association, Croatian Library Association, and Croatian Museum Association, each of them taking turns as the main organiser. The supporting co-organisers are Croatian State Archives, the National and University Library in Zagreb, the Museum Documentation Centre and the departments for Information and Communication Sciences of the universities of Zagreb, Zadar, and Osijek. In the first years of the seminar, Srce (University of Zagreb, University Computing Centre) and CARNet (Croatian Academic and Research Network) supported the seminars with technical and technological expertise, while ICARUS Croatia joined as a co-organiser for the 23<sup>rd</sup> seminar. The seminars were financially supported by the Ministry of Science and Education in the first couple of years, while the Ministry of Culture, now the Ministry of Culture and Media, has been supporting it throughout. The Ministry of Culture and Media has also provided financial support for the publication of the proceedings. Thus far, 24 proceedings from 24 seminars have been published, serving as evidence and as an archive documenting the progress, development of ideas, and accomplishments in the field. A particularly important result of the joint project of all four communities - library, archive, museum, and academic - is the publication of the Rules for Description and Access to Resources in Libraries, Archives and Museums in 2021 by the National and University Library in Zagreb, the Museum Documentation Centre and the Croatian State Archives (Vukadin 2021).

The seminars are organised within the structure of paper presentations, workshops and poster sessions, with occasional panel discussions on the topic of the annual seminar. The 13<sup>th</sup> AKM seminar (2009) was the first one to be organised on a specific topic. The topic was "Back to the Future: Revenge of the Standards!" as an expression of the necessity to get back to the roots, that is, to revisit the topic of standards, and their function and use in LAM information systems and projects. Topics of the following seminars evolved out of the current ones; some of them were: "Information Landscape Design"; "Object or Collection: What Is It We Are Documenting?"; "Physical or Virtual"; "Object, Space, Time"; "Global and Local, and Local and Global – GLOCAL"; "Common but Different: Principles of Interoperability Within LAM Community".

Each seminar opens with theoretical issues: the theory of information sciences as such, research of and revisiting the core elements like what is the object of interest (analogue and digital), that is, what has been chosen as the object for description and curation, then, ethics of librarianship, the mission of the museum, archival theory and technology, user studies, what, why, how and for whom to digitise, education in LAM community. The next block of topics covers technical issues of information organisation: description, analysis, and questioning of current content and technological principles, models and standards, and their application in information systems, user services and projects enabled and supported by information infrastructure like the internet, WWW and the Semantic Web; interoperability issues, that is, possibilities of exchange and re-use of data between two or more systems/ communities: the culture of interoperability, designing efficient and economically viable systems, and sustainable digital projects and portals that would eventually become a one-stop place for users who are interested in getting answers to a particular question regardless of which institution curates the object of their interest. The third block is dedicated to projects and pilot studies developed around a particular topic. The seminar's chosen topic is further detailed in workshops, which also cover complementary topics on conservation and preservation of LAM objects, museum and archives user education, users of information institutions (their needs and how to identify them), tourism (creative space of communication, publishers, and LAM community), etc.

## ON THE RELEVANCE OF POLITICS AND NORMS UNDERPINNING INFORMATION-FINDING AIDS

In the process of cross-fertilisation between and among LAM communities, "historical and contemporary understandings and manifestations of concepts authority, provenance, authenticity, evidence" (Willer 2018) have been recognised as highly relevant for the curatorial and presentation aspects of the cultural heritage held in these institutions. It has become necessary that these concepts, often understood as primary archival ones but implicitly embedded in cataloguing and curatorial traditions, come to the foreground of our dealings with the object of description and the description itself on both theoretical and practical levels. Even more so as the contemporary and future information environment built on linked open data, i.e., RDF statement<sup>11</sup> (W3C. Resource Description Framework) and Semantic Web technologies are designed on two operational principles – AAA and OWA:

The AAA principle, or "Anyone can say Anything about Anything", consequence: there always can be something else that somebody can say. This implies that the "truth" of an RDF statement cannot be assumed. There is no intrinsic way of detecting deliberate misinformation or metadata of very poor quality in an RDF statement. The Open World Assumption (OWA), a consequence of the AAA principle, means that at any time some new information can appear. This implies that none of the entities have complete descriptions, and it is assumed that new statements can always be made about any thing of interest. The RDF graph of linked data is always growing, and it is a simple process to add any new information about a resource (Dunsire 2020: 47).

In this context, Berry and Fagerjord were indicative in explicating the relevance of information retrieval (IR) tools while arguing that digital humanities should "focus on the need to (...) foreground the importance of the politics and norms that are embedded in digital technology, algorithms and software" (Berry 2017: 137). The politics and norms which libraries, archives, and museums are embedding in their IR tools – specifically in the design of user finding aids enabled by digital technologies, are expressions of "diverse cultural, community, disciplinary, professional and technological contexts, as well as (...) the nature, valence and relevance of the (...) concepts [authority, provenance, authenticity, evidence]" (Willer 2018: 8–9).

<sup>&</sup>lt;sup>11</sup> Resource Description Framework (RDF) is a standard model for data interchange on the web. RDF defines classes (entities) and properties (attributes and relationships) in a triple syntax: subject (class, domain of a triple) – predicate (property) – object (range of a triple: literal/vocabulary value or another class): RDF specifies that all three parts of the statement can be identified by an IRI (Internationalised Resource Identifier), while the object can also be a literal (W3C. RDF). For more detail see Willer and Dunsire 2013.

#### IVO MAROEVIĆ ON SCIENTIFIC AND CULTURAL INFORMATION

That the idea of the relevance of information-finding aids is not a new one but needs revisiting, especially in times of technological change which begs, in turn, for the change of politics and norms, can be read in Professor Ivo Maroević's rethinking of the concept of "cultural information"<sup>12</sup> in the LAM context in his keynote opening of the 2<sup>nd</sup> AKM seminar in 1998. In the keynote entitled "Problems of Identification and Description of Cultural Information (In Museums, Libraries and Archives)", Maroević, following Miro Tuđman, lays out a distinction between two kinds of information recognised in the theory of information sciences. These are: scientific and cultural information. Scientific information is

analytical and precise. It deals with the detail consisting of usually static standardized data whose aim is to maximally precisely *identify* the topic of the material. Although the data differ depending on the type of resource and institution, its import into information systems produces enormous information resources. However, its fundamental disadvantage is in its increase in quantity and therefore decrease in precision of choice, although making it appear that the precision of choice is high (Maroević 2000: 18).

Structural or cultural information

is rare because it requires concrete rendering of the cultural message. It comes down to the identification of the *secondary meaning* of the object of material. It deals with value, reason, meaning or necessity. Exactly because of that it depends on the subjective evaluation of the person who produces such an information, but also expresses the spirit of the time of their creation. In other words, cultural information accompanies the lifecycle of the object and represents its social relevance or, (...) its historical identity (Maroević 2000: 18–19).

When formalised, Maroević draws his conclusion, such "cultural information will become data and will be converted into a stable document that will transfer in time the testimonies of some other times in relation to the material that we had identified as cultural good" (Maroević 2000: 22–23). Irena Kolbas, in her presentation at the following, 3<sup>rd</sup> AKM

<sup>&</sup>lt;sup>12</sup> Maroević mentions scientific and cultural information carried by a museum object in his Uvod u muzeologiju (Introduction to Museology) (1993: 218).

seminar rightly pinpointed that view naming LAM institutions "factories of identity" (Kolbas 2000).

# IFLA ISBD FOR MANIFESTATION: TRANSCRIBED AND RECORDED DATA

The requirements put on information organisation politics and norms in order to distinguish yet complement "scientific, analytic and precise" data and "structural, cultural" data of "social relevance" in a finding aid, could be recognised as concepts on which the new conceptual Library Reference Model by the International Federation of Library Associations and Institutions (IFLA LRM) (Riva 2017) was designed. In order to describe how these concepts have become explicit in an implementation of the model, we will present here the current work on the revision of the IFLA standard for bibliographic description – ISBD: International Standard Bibliographic Description (ISBD 2011).<sup>13</sup>

The ISBD standard, designed at the beginning of the 1970s for the then-current technology - printed card catalogue, since then underwent a series of updates in order to adopt and integrate specificities of different types of library resources and modes of their issuance: analogue and digital. Thanks to that standard, MARC (MAchine Readable Cataloguing) formats have been developed, and today we have widely used and highly functional national and international catalogues and data-sharing systems (Gorman 2014). However, the context of the completely changed paradigm of approaching the bibliographic universe, expressed by the newly adopted conceptual IFLA LRM based on models and technologies of the Semantic Web, required a radical restructuring of the bibliographic standard. The IFLA ISBD Review Group, the group responsible for its maintenance, established the ISBD Manifestation Task Force in 2019 and charged it with the two-phase task to meet that requirement. These are: (1) develop ISBD element set and stipulations to describe Manifestation, and (2) produce a working document reflecting on transforming the ISBD into a full implementation of the LRM (Willer 2021).

<sup>&</sup>lt;sup>13</sup> Update to the ISBD 2011 edition was published in 2021.

IFLA LRM conceptual model, designed within an enhanced entityrelationship modelling framework, recognises four basic entities, i.e., points of user's interest: Work (the intellectual or artistic content of a distinct creation), Expression (a distinct combination of signs conveying intellectual or artistic content), Manifestation (a set of all carriers that are assumed to share the same characteristics as to intellectual or artistic content and aspects of physical form), and Item (an object or objects carrying signs intended to convey intellectual or artistic content). These entities can develop different types of relationships with entities Agent (an entity capable of deliberate actions, of being granted rights, and of being held accountable for its actions), Place and Time-Span, and are associated with entities Res (any entity in the universe of discourse) and Nomen (an association between an entity and a designation that refers to it) (Riva 2017). Entity Manifestation was chosen as the focal entity for the development of the new standard because traditionally ISBD is primarily focused on the description of a particular edition, that is, Manifestation that "represents the common characteristics shared by those carriers, in respect to both intellectual or artistic content and physical form" (Riva 2017: 25), with minor coverage of other three entities - Work, Expression, and Item.

As one of the basic principles of the ISBD for Manifestation (ISBDM), the Task Force recognised the distinction between transcription and recording of data as a two-part description process. Namely, the current ISBD stipulates elements and their data values as part of a chain of interdependent data forming a "record" in which descriptive data that is transcribed from a resource is intersected with data from sources external to that resource, and without forming explicit relationships between entities that make its aspects - Work, Expression, Manifestation, and Item. The technology of (Semantic) Web environment, informed by the AAA and OWA principles, and RDF model, requires such an explicit statement: each set of data (attribute, e.g. title) should be uniquely identified by pertaining to a particular entity (a domain, e.g. Manifestation), and, if forming a relationship between entities, for example, between manifestation and its creator, that relationship should be uniquely identified (domain: manifestation, relationship: creator agent of manifestation, range: "Fraktura").

ISBDM stipulates rules for transcribing a manifestation statement, that is, data is transcribed from the manifestation - an object of description, in form and content as it is found there, without any intervention from the part of the cataloguer. In this respect, the description follows the principle of Representation of the IFLA's Statement of International Cataloguing Principles (ICP): "A description should represent a resource as it appears" (IFLA 2017: 5). It would support the LRM user task Identify manifestation, "[t]o clearly understand the nature of the resources found and to distinguish between similar resources" (Riva 2017: 15). It would consist of, so-called, literals, string of signs (text, numbers) to be displayed to the user in an information retrieval system. Such data being very precise in itself is useful for keyword indexing to support the user's task to find the manifestation, but being uncontrolled by a cataloguer would have high recall and low precision. This is recognised by Maroević: "scientific information" or data whose quantity eventually decreases the precision of the system as a whole, i.e. much of the retrieved information is not relevant. Such an aspect of transcribed or "scientific" data is also argued by Svenonius: "Accuracy in description does not by itself ensure effective retrieval. (...) The principle of representation is useful in developing means to achieve the finding and identifying objectives, but here its usefulness stops" (Svenonius 2000: 75).

Recording means that data is the result of the cataloguer's "intervention" providing it in this process with "authority, provenance, authenticity, evidence", and thus making it "culturally relevant". The recording is a complex process, requiring the high expertise of a cataloguer in describing the object of interest by applying particular cataloguing standards. The data produced would be expected to meet additional LRM user tasks which refer to the entities and their properties:

- to find and explore the context of entity relationships find: "To bring together information about one or more resources of interest by searching on any relevant criteria", and explore: "To discover resources using the relationships between them and thus place the resources in a context";
- to identify, select and obtain the entity by support of entity properties select: "To determine the suitability of the resources found, and to enable to either accept or reject specific resources" and obtain: "To access the content of the resource" (Riva 2017: 15).

In the context of declared IFLA's Statement of International Cataloguing Principles (ICP) the recorded data supports the general principles of "accuracy", "sufficiency and necessity", "significance", "economy", "consistency and standardization", "integration", and "interoperability" (IFLA 2017: 5–6).

The recorded data represents "things", entities of user interest, or nodes of the linked data graph enabling the user primarily to navigate the context of entity of interest in the information universe. This type of data provides high precision and low recall, and by its characteristics corresponds to Maroević's "cultural information" or data. This data is the product of a process of normalisation/authorisation according to locally implemented international standards which make them contextual to a specific culture expressing the "social relevance" of the object of description. On the example of the choice and form of author's name in the case of multiple identities made by different national bibliographic agencies, Dunsire and Willer are forwarding the question of "cultural message" of data to the general question of "how to make explicit the social constructs, the rules that govern and underpin data production for consumption by the human user, to the "social machines", the rules that govern and underpin data production for machine processing and semantic reasoning in applications aimed at the human user" (Dunsire and Willer 2018: 99-100). They conclude that "Semantic Web technologies support this approach [of semantic coherence in the categorisation of entities and relationships of interest], especially in the context of social construction and deconstruction. The context of the AAA principle and Open World Assumption implies an increasing importance of the role of official, national cultural heritage organisations as providers of trustworthy provenance" (Dunsire and Willer 2018: 111).

#### CONCLUSION

We can argue with certainty that Maroević's concepts of the complementary "scientific information" or data and "cultural information" or data are relevant today, as they had been relevant when expounded. The researcher in the digital humanities would need both types of data, as they need to identify the object of interest, but also need to explore and have available access to the cultural context of that object which is provided by the cataloguer or curator within the scope of their mission. The Semantic Web technologies' operational principles – AAA and OWA provide opportunities and means to revisit and enrich Maroević's concept of cultural information and transform it from a fixed document into a dynamic one "that will [continue to] transfer in time the testimonies of some other times in relation to the material that we had identified as cultural good" (Maroević 2000: 22–23).

The notion of the continuity of rethinking the theoretical background behind principles and models that govern the design of standards and information-finding aids used in a particular cultural environment and at a particular point in time, and of recognition of their interdependence with current information technologies is being confirmed.

#### **REFERENCES LIST**

- Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture. Available at: https://akm.hkdrustvo.hr/ (accessed 28 March 2023).
- Berry, David M. and Anders Fagerjord. 2017. Digital Humanities: Knowledge and Critique in a Digital Age. Cambridge, England: Polity Press.
- Dunsire, Gordon. 2020. "Telling Tails: Metadata Standards and the Digital Humanities". In Empowering the Visibility of Croatian Cultural Heritage through the Digital Humanities. Marijana Tomić, Mirna Willer, and Nives Tomašević, eds. Newcastle upon Tyne: Cambridge Scholars Publishing, 40–52 [definitions edited by G. Dunsire].
- Dunsire, Gordon and Mirna Willer. 2018. "Authority versus Authenticity: The Shift from Labels to Identifiers". In Authority, Provenance, Authenticity, Evidence: Selected Papers from the Conference and School Authority, Provenance, Authenticity, Evidence, Zadar, Croatia, October 2016. Mirna Willer, Anne J. Gilliland, and Marijana Tomić, eds. Zadar: Sveučilište u Zadru, 87-115. Available at: https://morepress.unizd.hr/books/index.php/press/ catalog/view/46/42/595 (accessed 28 March 2023).
- Gorman, Michael. 2014. "The Origins and Making of the ISBD: A Personal History, 1966–1978". *Cataloging & Classification Quarterly* 52: 821–834. Available at: https://doi.org/10.1080/01639374.2014.929604 (accessed 28 March 2023).
- IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code. 2017. Statement of International Cataloguing Principles (ICP). 2016 Edition with minor revisions, 2017 by Agnese Galeffi (chair), María Violeta Bertolini, Robert L. Bothmann, Elena Escolano Rodríguez,

and Dorothy McGarry. Available at: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp\_2016-en.pdf (accessed 28 March 2023).

- ISBD International Standard Bibliographic Description. Consolidated edition 2011. Den Haag: International Federation of Library Associations and Institutions. Available at: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/isbd/isbd-cons\_20110321.pdf (accessed 28 March 2023).
- ISBD International Standard Bibliographic Description. 2021 Update to the 2011 Consolidated Edition. Available at: https://repository.ifla.org/handle/123456789/1939 (accessed 28 March 2023).
- Kolbas, Irena. 2000. "Tvornice identiteta". In 2. i 3. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova. Mirna Willer and Tinka Katić, eds. Zagreb: Hrvatsko muzejsko društvo, 99–106 [citation translated into English by M. Willer].
- Maroević, Ivo. 2000. "Problemi identifikacije i obrade kulturnih informacija (u muzejima, knjižnicama i arhivima)". In 2. i 3. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova. Mirna Willer and Tinka Katić, eds. Zagreb: Hrvatsko muzejsko društvo, 17–23 [citations translated into English by M. Willer].
- Maroević, Ivo. 1993. *Uvod u muzeologiju*. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu, Zavod za informacijske studije Odsjeka za informacijske znanost.
- Riva, Pat, Patrick Le Bœuf, and Maja Žumer. 2017. IFLA Library Reference Model: a Conceptual Model for Bibliographic Information. Den Haag: International Federation of Library Associations and Institutions. Available at: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017\_rev201712.pdf (accessed 28 March 2023).
- Svenonius, Elaine. 2000. The Intellectual Foundation of Information Organization. Cambridge, London: The MIT Press.
- Vukadin, Ana, ed. 2021. Pravilnik za opis i pristup građi u knjižnicama, arhivima i muzejima. Zagreb: Hrvatski državni arhiv, Muzejski dokumentacijski centar, Nacionalna i sveučilišna knjižnica u Zagrebu. Available at: https:// pravilnik.kam.hr (accessed 28 March 2023).
- Willer, Mirna. 2021. "Implementacija LRM-a u ISBD-e: od IFLA-ina knjižničnog referentnog modela (LRM) do novoga Međunarodnog standardnog bibliografskog opisa (ISBD)". In 24 seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova. Tatjana Šarić and Mirna Willer, eds. Zagreb: Hrvatsko arhivističko društvo, 105–116.
- Willer, Mirna and Gordon Dunsire. 2013. Bibliographic Information Organization in the Semantic Web. Oxford: Chandos.

- Willer, Mirna, Anne J. Gilliland, and Marijana Tomić, eds. 2018. Authority, Provenance, Authenticity, Evidence: Selected Papers from the Conference and School Authority, Provenance, Authenticity, Evidence, Zadar, Croatia, October 2016. Zadar: Sveučilište u Zadru. Available at: https://morepress. unizd.hr/books/press/catalog/book/46 (accessed 28 March 2023).
- W3C. Resource Description Framework (RDF). Available at: https://www.w3.org/RDF/ (accessed 28 March 2023).
## ANALYSING ONLINE HORROR FICTION WITH CORPUS LINGUISTIC TECHNIQUES FOR ENGAGING MUSEUM LABELS: CREEPYPASTA WIKI AND THE HOUSE OF TERROR'S MOBILE APPLICATION

#### DANIEL IHRMARK

Linnaeus University Växjö, Sweden daniel.o.sundberg@lnu.se https://orcid.org/0000-0002-0930-644X

#### ZOI TSIVILTIDOU

University of Leicester Leicester, England https://orcid.org/0000-0003-3526-6602

This paper explores how the language and content of a collection of born-digital horror narratives could potentially inform the writing of museum labels in horror-related exhibitions. 155 stories from Creepypasta Wiki were selected based on user ranking with a high rating being considered as indicative of engaging and immersive stories; they were then compared to 79 labels from the mobile application of the House of Terror museum in Hungary. The analysis focused on categorisation and frequent nouns and pronouns. The results indicate that first-person narratives and relatable settings are often used to create immersion and engagement in the horror narratives, while the museum labels focus on large-scale collective actions from a more distant perspective.

Keywords: online horror fiction, museum labels, House of Terror, mobile application, engagement, text analysis

#### **1. INTRODUCTION**

This paper looks into how the common language conventions in online horror fiction could potentially inform the writing of museum labels in a mobile application, as in the House of Terror in Hungary.<sup>14</sup> The aim is to explore the possibility of building a connection between labelling for a horror-related exhibition and user preferences in online horror fiction. The stories in the Creepypasta Wiki<sup>15</sup> offer a valuable resource for finding what kind of content and descriptions commonly appear in toprated stories, suggesting that certain stylistic traits lead to immersion and engagement. 155 stories from Creepypasta Wiki and 79 labels from the mobile application of the House of Terror were analysed comparatively in order to highlight differences. The understanding of different language patterns and their implications in inducing reader responses can theoretically guide how difficult (hi)stories are told. The insights point towards the possibility of the users' preferences in horror narratives informing the writing of engaging museum labels in the future.

#### 2. CONTEXT

#### 2.1. CREEPYPASTA WIKI AND USERS' PREFERENCES

While discussing the contents of Gothic horror, Bloom connects it to the needs of the reader and the dark side of experienced domestic life; a genre "obsessionally connected with contemporary fears" (1998: 2). In *Gothic Romanced* (2008), Botting traces the development of both Gothic horror and Romantic fiction into the modern age, detailing the Gothic influences on genres popular today to show how the connection to societal fears disseminate into the broader genre of horror. In a later text, *Limits of Horror* (2013), Botting carries this over to the digital medium and discusses how this influence is continued on-screen through new modes of media.

"Creepypasta" is a portmanteau formed by "creepy", indicating horror as the genre, and "pasta" derived from "copypasta", which is a form

<sup>&</sup>lt;sup>14</sup> https://www.terrorhaza.hu/en.

<sup>&</sup>lt;sup>15</sup> https://creepypasta.fandom.com/wiki/Creepypasta\_Wiki.

of online fiction produced in a shape fitting for posting on forums and image boards (Blank and McNeill 2018). This distribution model also enabled a preference for faux-reality in combination with first-person narratives, the former often made possible due to the virtual anonymity of the author (Williams 2015, Blank 2007). Many of the stories draw on urban legends and folklore in their relationship to the real world, leading to the concept being discussed as "Netlore" in previous research (Sanchez 2019).

Both as an inheritance from previous horror fiction and as a part of the new focus of the digital setting, Creepypastas rely heavily on engagement and immersion to drive the story. This plays into the form of the story, commonly framed as the author having a conversation with the reader online, immersing the reader in the faux-reality of the story by drawing on the already asynchronous exchange and anonymity of the internet's text forums.

#### 2.2. STORYTELLING IN MUSEUM LABELS

Stories in museums have been widely used to bring objects to life, representing and interpreting content for curatorial purposes "for museums are storytellers" (Bedford 2001: 33). Curatorial stories were authentic, reliable exhibition backdrops for each theme to unfold. Museums told stories through the collection, curation and meaningful display of artefacts and the use of explanatory visual and textual content (Roussou 2001: 185).

Labelling played a critical role in informing visitors about people, places, events and ideologies, and communicating certain messages (Bitgood 2000; Screven 1992; Serrell 1996). Captions are often limited to 50-60 words, so-called "tombstone labels". Tombstone labels provide the most basic information about the object on display, usually the name, age, place of manufacture, museum owning it, and catalogue number. Interpretive labels talk more about an object than identification labels. These are usually written in full sentences and often include a heading.

Caulton claimed that "[t]here is a purpose behind each display, a story to tell with each exhibit, an idea to unfold in each gallery" (1998:

5). Objects told stories about their owner(s), function, and context, and through these visitors were led to different associations depending on how the stories were told and interpreted (Dudley et al. 2012). Storytelling allowed visitors "to connect to these artefacts on a deeper, more personal level" (Springer et al. 2004: 3) and increased accessibility (Pujol-Tost et al. 2013). Nonetheless, pre-set stories were renegotiated and reconfigured as subjective, fixed carriers of statements. Runnel and Pruulmann-Vengerfeldt (2012) identified, thus, the risk of marginalising or emphasising some interpretations over others.

Recently, more and more museums ask visitors to engage with the content emotionally and intellectually, and "gain meaningful understanding of museum objects" (Kilfeather 2013: 106). Storytelling is no longer a monolithic output of the museum's written communication strategy, but rather a dialectic process that involves the visitors as much as the curators. The stories museum labels tell are now seen as "a potent way for museums to present inclusive and nuanced history, to make big ideas less overwhelming and abstract, and to create frames of experience that encourage deep and satisfying engagement for visitors" (Wong 2015: 2). And this is exactly the focus of this research: to question whether user preferences could initiate a dialogue on how engaging museum labels are written and how difficult (hi)stories are told.

#### 2.3. HORROR-RELATED EXHIBITIONS

Human vulnerability in museology has been studied in terms of sociopolitical implications (Silvén and Björklund 2006; Sandell 2007; Tinning 2017; Violi 2012); but in linguistic terms, it has not yet been examined for observed, or predicted, associations prompted by engaging labelling. Different terms have been used to address sensitive themes: some are "difficult exhibitions" (Bonnell and Simon 2007), "difficult heritage" (Macdonald 2009; Witcomb 2013), "hot or taboo topics" (Williams 2007), "contentious topics" (Cameron and Kelly 2010), "difficult knowledge" (Lehrer et al. 2011; Simon 2011), "controversial topics" (Tøndborg 2013), "challenging history" (Kidd 2014) or "difficult history" (Rose 2016).

An exhibition is considered difficult if its content is experienced as eliciting the burden of "negative emotions" (e.g., shame, guilt, melancholia or hatred) often associated with a departure from positive pursuits (Bonnell and Simon 2007: 67). Such exhibitions fall within the realm of dark tourism studies (Lennon and Foley 2000), which relate to "difficult heritage" and memory discourse.

In the last three decades, dark tourism research has addressed terror- and horror-related events in sites associated with death, suffering and difficult history (Roberts and Stone 2014); and it has sought out visitor motivations (Tarlow 2005), which vary "from morbid fascination or 'rubber-necking', through empathy with the victims, to the need for a sense of survival/continuation, untested factors which, arguably, demand verification within a psychological context" (Stone 2008: 579).

In our research, horror-related content is considered any textual information that includes direct or implicit references to traumatic experiences and sensitive themes. "An experience is traumatic if it is overwhelming enough to cause a break in the linkage and meaning of experience, in narrative memory" (Howell 2020: 31). An exhibition about that does more than acknowledge trauma; it recognises it and through storytelling, it connects visitors to a shared humanity, fosters cultural humbleness, and refrains from othering. Labels are expected to acknowledge and contextualise difficult (hi)stories, involving the relief of suffering, validation for what has been endured, and words that seek to make sense out of something horrific.

This means that "creating effective labels is not just a matter of generating lots of attention and effort, but [it] depends on making mindful attention more fun than random or causal manipulations" (Screven 1992: 190). Labels should not only communicate messages and answer questions, but they should also ask questions and prompt visitors to connect unfamiliar objects or themes to familiar experiences. Research has shown that "visitors *do* read labels if they perceive that labels will meaningfully help relate exhibit content to them or will provide feedback and follow-up to exhibit experiences" (Screven 1992: 208, italics in original). Besides format and context, language use affects how attention is attracted and how a response is elicited. In that light, museum labels should speak "the language of today and tomorrow" (Meereboer 2018: 34) and "the information provided by labels should contain information visitors want to know" (Screven 1992: 184).

#### 3. METHODOLOGY

#### 3.1. OBJECTIVES

The purpose of the study was to identify similarities and differences in language conventions in online horror fiction and the mobile application of the House of Terror and, therefore, to explore a theoretical connection between user preferences and the text that is selected for museum labels of difficult heritage exhibitions. This applies to what kind of stories museums tell, and what kind of visitor engagement they aim for based on language conventions, as a theoretical formulation of written communication and as a proposal for re-examining the language that is used. The House of Terror was chosen due to it fitting the context and scope of our research. The building itself was used as the headquarters of Hungary's Arrow Cross Nazi party. It houses a permanent exhibition on the Hungarian totalitarian dictatorships across four floors with 28 rooms, which is accessible via the Internet and through a mobile application (available in Google Playstore © for Androids). The application offers a recommended tour titled "Explore the House of Terror Museum", with 52 exhibits. The user interface displays photographs of the rooms, maps, and photographs of the objects that are accompanied by labels and interpretative text, including direct quotes. The museum's mission is to make visitors understand that the sacrifices, horrific acts, and fights for freedom and independence in the 20th century were not in vain. The terrors described within this exhibition are situated historically, starting from the period under Nazi rule and the emergence of the Arrow Cross party, moving on to the taking over of communist rule after World War II, when Hungary was part of the Eastern Bloc under Soviet dominance, and finishing with the revolution of 1956 that was crushed by the Soviet Red Army.

#### 3.2. DATA COLLECTION

Queries were carried out in two corpora as shown in Table 1, harvested from Creepypasta Wiki and the mobile application of the House of Terror from June to August of 2019 using a BeautifulSoup web-scraper written in Python 3 ©.<sup>16</sup>

<sup>&</sup>lt;sup>16</sup> Both corpora are available upon request.

Corpus:	Word count:	No. of texts:
CreepyCorpus	454,960	155
HoTCorpus	8,464	79

Table 1: The datasets.

The analysis was performed using the AntConc ©, Voyant © and NLTK © libraries. Some visualisations and pre-processing were carried out using Orange 3 © (see appendix B for tools). Three aspects were explored: 1. category distribution (CreepyCorpus only), 2. content, and 3. story perspective.

In the CreepyCorpus, the inclusion criteria were original content for the medium that appeared in the top-rated section for each of the 43 categories available on the site (Appendix A). Looking at the category distribution in the dataset containing duplicate stories provided information on which categories were found engaging. The content analysis of the non-duplicate dataset initially made use of a visualisation of the texts with the stop words removed. We then moved on to focusing more precisely on nouns and pronouns. By looking at the commonly used pronouns, we intended to provide insight into the preferred narration.

The TreeTagger © tag set in combination with Lawrence Anthony's TagAnt © interface was used to part-of-speech (POS) tag the texts, which specify the function and word class of every word in the stories. This allowed for specified queries into word classes such as nouns and pronouns. The TreeTagger used is shown to have a 94% accuracy when dealing with written language, making it reliable enough for this study, as it is mainly focused on overviews (Horsmann et al. 2015). The same analysis was then performed in the HoTCorpus.

#### 3.3. LIMITATIONS

The datasets used for this study are small, and no general conclusions regarding the broader genre of online horror fiction should be drawn from our analysis due to data sparsity. However, due to the CreepyCorpus including only the top-rated stories from each category, the content should be representative of material found popular by the Creepypasta Wiki's users. As it is this positive audience engagement this study intends to explore, the dataset was considered viable. The observed preferences could thus potentially inform language used for engaging museum labels, given that other considerations regarding the function of labels are reflected upon.

Nonetheless, empirical research of the practical implications should follow to examine in situ the impact and value of adjusting the text of the museum labels based on user preferences. Further research is expected to yield interesting results in what language conventions visitors report as engaging and how audiences relate to the text and, therefore, to the objects in an exhibition through the text itself.

#### 4. FINDINGS

# 4.1. COMMON LANGUAGE CONVENTIONS IN THE CREEPYCORPUS

The category distribution provided broader strokes of content trends within the dataset. It is important to note that the stories have been assigned categories by the archive's users and not by the authors of this article. Each of the 43 category headings also contains several sub-categories.



Figure 1: Category distribution within the Creepypasta Wiki archive.

As seen in Figure 1, the top two counts in the category dataset seem to reflect rather traditional aspects of horror fiction: "mental illness" and "history". The description of the genres from the archive reads "Mental illness: Pastas involving various levels of insanity" and "History: Pastas involving historic events", both very broad categorisations. This is not the reason for the high counts, though, as there are much larger categories that rank lower in terms of texts within the top-ranked stories (see Appendix A).

In the non-duplicate dataset, initial content analysis is visualised as a word cloud in Figure 2 and provides an overview of what is found in the dataset. Sizes are relative to the frequency of the word.



Figure 2: Initial language overview.

The high frequency of the auxiliary verb "could" suggests the past tense is a common feature in the stories. Furthermore, as a modal verb, it gives some insight into the possible mode of the stories; i.e., possibility or uncertainty in this case. "Time" is also of interest, as a collocate query shows it occurring together with "this", "at the", and "by the" at high frequency, as "this time" indicates repeated encounters or milieus within the stories, while "at the time" or "by the time" indicate reflective narration. However, the latter two could also be applied stylistically, especially "by the time" as a marker of urgency. Looking at the nouns revealed what types of content the stories contain and gave us a better idea of the "what" and "where" of the stories. The nouns were extracted from the non-duplicate dataset using the tags (\*\_NN or \*\_NNS).



Figure 3: Frequent nouns in the Creepypasta Wiki corpus.

"Time" (1130), "room" (745), "door" (735), "man" (671) and "way" (665) make up the five top frequency nouns in the dataset, as seen in Figure 3. "Time" has already been expanded upon, but "room" provides an idea of the setting for the stories, linking well together with the frequent use of "door". Figure 4 provides KWIC (Key Word In Context) lines for a non-sequential sample of "room".

Figure 4 shows a varied context for "room" and a collocate query was performed to further understand the use of the environment. The query was performed on the first word appearing in the position to the left of "room"; the L1 collocates. The collocates are listed to the right in Table 2 below, while the frequency of occurrence in the dataset is listed to the left.

room . And , now , there was Alice '
room and onto the altar . I
room and opened the top right
room, and perhaps one of them
room and play video games . He
room and pressed A but nothing
room and sat on the couch .
room and shut the door on
room and slammed the door shut
room and slammed the door . I
room and slammed the door . I
room and started walking down yet

Figure 4: "Room" in the context of a non-sequential sample.

Frequency	Collocate
163	the
39	my
35	living
16	his
14	that
12	а
9	portrait
9	door
7	this
7	dining

Table 2: L1 collocates of "room".

Among the collocates, two possessive pronouns are found: "my" and "his". This indicates the "room" is often defined as a personal space for the characters within the story, either as the narrator's or belonging to another character. At ranks 3 and 10 "living" room and "dining" room are found. This goes further to indicate the home or living quarters as a common setting.

"Man" followed "room" and "door" at a frequency of 671. Among the top L1 collocates both the definite and indefinite articles are found, but beyond the function words some descriptors are also found: "old", "young", "slender", "tall", "goat", "hooded", "younger", "little", "large", "elderly", and "dead". Most of the L1 collocates to "man" roughly fall within the categories of age and physical appearance. Outliers would be "goat", "hooded", and "dead". However, they all appear in low, singledigit frequencies.

"Way" is found as "the way" and "my way" at the top of the frequency list of L1 collocates, with "his way" and "its way" following in 4<sup>th</sup> and 5<sup>th</sup> place. In 3<sup>rd</sup> place, however, "no way" is found. Out of 35 occurrences of "no way", 14 are found in the structure "there was no way". Within the list, 4 are variants of "there was no way \*pronoun\* could" and 3 "there was no way out". This indicates a loss of agency and uncertainty about one's ability to handle the current situation. Such insight is useful in contextualising an object within a storyline of vulnerability and distress, as a particular choice of words may intensify the message more than another.

The noun overview showed rooms as a frequently occurring space, with collocates indicating accommodation as the environment by the use of "living room" and "dining room". This is relevant to how museum labels could theoretically "build" a sense of place within the exhibition space and how reference to an enclosed, indoor milieu, particularly to an intimate and familiar room, could intensify emotional connections with a display and contribute to a bodily experience by evoking personal memories or framing scenarios.

Also, we focused on the singular pronouns because we were interested in the narrator's voice. The strong preference for "I" indicates that the first-person narration is the most prevalent in the dataset, while the third-person also occurs quite frequently (Figure 5). "It" was included due to the rather high frequency of non-human characters involved in certain categories, such as monsters, ghosts and other creatures. However, due to "it" also referring to objects, the query returns little that is of value. Interestingly, "you" is found at a comparatively high frequency, indicating usage beyond that in dialogues.



Figure 5: Pronoun Distribution in CreepyCorpus.

In order to expand on the use of "you", a KWIC line query similar to the one used for "room" was carried out. It appears that it is used as an interaction with the audience, either at the beginning or the end of a story.

The pronoun distribution shows the first-person singular as the most common, indicating a preference for first-person narration. The identity of the narrator often seems to be a child or young person, although the context often indicates the stories being relayed by an older version of the narrator. The preference could also be linked to the immersive intent of the stories, as the faux-reality upheld through the formatting of stories as diaries, forum posts or similar demands a sequence of events where the author is describing a situation that took place before the author wrote it down.

The preference for first-person narratives points to a framing of the interpreting experience that not only supplies the frame through the narrator but also includes reflective prompts for the reader to participate in the story. This was shown to be a feature mainly deployed by immersing the reader through direct involvement, i.e., using the "you" pronoun as a means of speaking to the reader directly. This aligns with what the literature about museum engagement suggests in terms of thought-provoking personalised interaction (Ardito et al. 2009; Arvanitis 2005; Bearman and Geber 2008). Such wording in a museum label could be used to "create frames of experience that encourage deep and satisfying engagement for visitors" by having the narrator act as a framing device for the exhibition (Wong 2015: 2).

A label in the mobile application could use the trajectory of the curator's story as a starting point for the visitor's reflective observation and a re-appropriation of the content. In order to engage visitors, museums are called upon to be "responding to the choices its users make" (Black 2012: 11). Different perspectives could come forward, reframing perceptions along the way. Visitors are demanding a greater say in interpretation, reflecting a growing recognition in museum practice and the wider society that museums are not neutral and benefit from a multiplicity of voices. Using labels to bring new and multiple voices to exhibitions is becoming more common as museums seek to become more inclusive, accessible, and democratic.

#### 4.2. COMMON LANGUAGE CONVENTIONS IN THE HOTCORPUS

As with the previous corpus, the initial step of analysis is a general overview of frequently occurring items in the texts. While the focus of the Creepypastas was immersion and audience engagement, the focus of the House of Terror's exhibition and the text in the mobile application is grounded in describing a specific aspect of the national history of Hungary.

The initial language overview showcases the main content of the corpus, which aligns with both the topic and context of the exhibition (Figure 6). One interesting thing to note here is the prevalence of collective nouns as shown in Figure 7, which is quite different from the trends seen in the Creepypastas. Note that the document segments refer to the collected texts in order of appearance in the exhibition, but that the seg-

mentation itself is applied by Voyant Tools in order to create a timeline of the materials based on the tokens in the corpus. The document segments do not correspond to the rooms of the exhibition, nor the individual documents, as they are not based on any external factors of the texts.



Figure 6: Initial language overview of the House of Terror application.



Figure 7: Trends of the top five most frequent words in the HoT dataset.

Looking at the trends of the five words with the highest frequency, the words appear in all sections of the exhibit, albeit at differing frequencies depending on the focus. Moving on to the nouns (Figure 8), the insights gained from the initial overview hold true.



Figure 8: Frequent nouns in the HoTCorpus.

Interestingly, after the collective nouns and territorial descriptors ("country", "state") expected from the initial overview, "camp" and "room" are found. By using AntConc to examine KWIC lines, the context becomes visible, along with the usage of the words in the dataset (Figure 9). While "camp" is seen used both referring to an army camp and to describe the labour camps where some of the atrocities described in the museum's exhibitions took place, "room" is found in both descriptions of environments related to the story function of the text within the exhibits and as a reference to the spatial properties of the exhibition itself.

became their password. The room presents a mine car blindfolded from one interrogation room to another. Based on

next exhibit in the room, or to discover the are displayed in the room. Some of them contain in the party. The room symbolizes that all of This is the only room that was preserved in the middle of the room, the raw material of cell, with only enough room to stand. The floor

Figure 9: KWIC lines for "camp" and "room".

Moving on to the pronouns, "it" is by far the most common, followed by "you" and "he". Based on this distribution, both the actors involved and the perspective given are very different in the two corpora.



Figure 10: Pronoun Distribution in HoTCorpus.

In addition, "they" appeared at a standardised frequency of 6.37 / 1,000 in HoTCorpus and 4.03 / 1,000 in CreepyCorpus. This makes sense when looking at frequent words and nouns, as many are collective nouns meriting "they" rather than the singular pronouns, while the emphasis on parties and nations would also account for the use of "it". The pronoun used in the HoTCorpus indicates a descriptive narrative focusing on the actions of parties, collectives and other groups rather than on the individual experiences of the events. The emphasis is on the collective rather than on the individual. One explanation could be that the subject is tied to the actions of large groups of people, such as parties and armies. It seems that any emphasis on or repeated references to individuals would overshadow the impact of such collective actions.

The first-person singular appears only 4 times in the HoTCorpus and one of the occurrences is an incorrect OCR reading of a numerical. The remaining three occurrences stem from short quotations providing accounts of the events as seen in the KWIC lines (Figure 11).

#### Imre Kovacs 1 "Last night I dreamt that the Germans k Father Andras Kun, 1945 2 " I have breathed fire and emerged from World War I on the losing side. my first major interrogation, I prayed for the Lord

Figure 11: KWIC lines for "I".

We see that the HoTCorpus' conventions are rooted in the very specific nature of the exhibitions. While this explains the language used, it also highlights how an awareness of the language conventions at play could allow the labelling to appeal to visitors by drawing on areas highlighted by the stories in the CreepyCorpus; by giving space to individual experiences and narratives based on first-person descriptions. Settings that relate to historical events examined through the lens of traumatic experiences could become the backdrop for personal stories, which explore these themes further and prompt visitors to situate themselves in the context.

Similarly, the examples of the use of first-person singular pronouns in the HoTCorpus indicate that personal stories are already used to some extent. However, we should note here that the majority of the text in this specific exhibition deals with the grander narrative related to the high frequency of collective nouns.

#### **5. CONCLUSION**

The preferences in the two corpora are quite clear and could theoretically form a foundation for the integration of first-person narratives in the labels of exhibitions about horror, difficult (hi)stories and human vulnerability. Such knowledge is relevant and useful for curators and museum professionals alike, who seem to have understood that it is not enough to use informative language because the visitors who will be reading the labels are diverse and rather demanding. Our analysis of content and descriptions in user-generated online horror fiction could potentially inform the writing of engaging labels, and, therefore, becomes a resource for generating labels with such intent.

Nonetheless, there are other considerations to be made, and further empirical research is needed into the benefits of writing museum labels for a mobile application based on user preferences, as found in borndigital collections of stories. Simply changing or adjusting the text in museum labels is not enough to initiate a dialogue on how visitors engage with sensitive topics and how they interpret objects and connect on both cognitive and emotional levels to the exhibition stories. What is required is an in-depth long-term examination of pre-visit and postvisit visitor engagement with difficult heritage exhibitions and user preferences in online horror fiction to compare the common language conventions and to measure the impact and value of having user preferences inform the text used in museum labels.

#### **REFERENCES LIST**

- Ardito, Carmelo, Paolo Buono, Maria Francesca Costabile, Rosa Lanzilotti, and Antonio Piccino. 2009. "Enabling Interactive Exploration of Cultural Heritage: An Experience of Designing Systems for Mobile Devices". Journal of Knowledge Technology and Policy 22 (1): 79–86.
- Arvanitis, Konstantinos. 2005. "Museums Outside Walls: Mobile Phones and the Museum in the Everyday". In Proceedings of the IADIS International Conference on Mobile Learning, Mobile Learning 2005. Pedro Isaias, Carmel Borg, Piet Kommers, and Philip Bonanno, eds. Qawra, Malta: IADIS, 251–255.
- Bearman, David and Kati Geber. 2008. "Transforming Cultural Heritage Institutions through New Media". Museum Management and Curatorship 23 (4): 385–399.
- Bedford, Leslie. 2001. "Storytelling: The Real Work of Museums". Curator: The Museum Journal 44 (1): 27–34.
- Bitgood, Stephen. 2000. "The Role of Attention in Designing Effective Interpretive Labels". Journal of Interpretation Research 5 (2): 31–45.
- Black, Graham. 2012. Transforming Museums in the Twenty-first Century. Oxon: Routledge.
- Bloom, Clive, ed. 1998. Gothic Horror: A Reader's Guide from Poe to King and Beyond. London: MacMillan Press.
- Bonnell, Jennifer and Roger I. Simon. 2007. "Difficult' Exhibitions and Intimate Encounters". *Museum and Society* 5 (2): 65–85.
- Botting, Fred. 2008. Gothic Romanced: Consumption, Gender and Technology in Contemporary Fictions. Hoboken: Taylor and Francis.
- Botting, Fred. 2013. Limits of Horror: Technology, Bodies, Gothic. Manchester: Manchester University Press.
- Cameron, Fiona and Lynda Kelly, eds. 2010. Hot Topics, Public Culture, Museums. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Caulton, Tim. 1998. Hands-On Exhibitions: Managing Interactive Museums and Science Centres. Oxon: Routledge.

- Dudley, Sandra H., Amy Jane Barnes, Jennifer Binnie, Julia Petrov, and Jennifer Walklate, eds. 2012. Narrating Objects, Collecting Stories. Oxon: Routledge.
- Horsmann, Tobias, Nicolai Erbs, and Torsten Zesch. 2015. "Fast or Accurate?
  A Comparative Evaluation of PoS Tagging Models". In GSCL 2015. International Conference of the German Society for Computational Linguistics and Language Technology. Proceedings of the Conference. Bernhard Fisseni, Bernhard Schröder, and Torsten Zesch, eds. Germany: University of Duisburg-Essen, 22–30.
- Howell, Elizabeth F. 2020. Trauma and Dissociation Informed Psychotherapy: Relational Healing and the Therapeutic Connection. New York: Norton Professional Books.
- Kidd, Jenny. 2014. "Introduction: Challenging History in the Museum". In Challenging History in the Museum: International Perspectives. Jenny Kidd, Sam Cairns, Alex Drago, and Amy Ryall, eds. London: Routledge, 1–17.
- Kilfeather, Eoin. 2013. "Visualising Museum Stories". In Electronic Visualisation and the Arts (EVA 2013). Kia Ng, Jonathan P. Bowen, and Sarah Mc-Daid, eds. London: EVA 2013, 106–112.
- Lehrer, Erica, Cynthia E. Milton, and Monica Eileen Patterson, eds. 2011. *Curating Difficult Knowledge: Violent Pasts in Public Places.* New York: Palgrave Macmillan.
- Lennon, J. John and Malcolm Foley. 2000. Dark Tourism: The Attraction of Death and Disaster. London: Continuum.
- Macdonald, Sharon. 2009. Difficult Heritage: Negotiating the Nazi Past in Nuremberg and Beyond. London: Routledge.
- Meereboer, Theodorus. 2018. "More or Less". In *Museums and Written Communication: Tradition and Innovation*. Ani Avagyan and Nick Winterbotham, eds. Newcastle upon Tyne: Cambridge Scholars Publishing, 13–35.
- Mitchell, Alex, Clara Fernández-Vara, and David Thue, eds. 2014. Interactive Storytelling. 7th International Conference on Interactive Digital Storytelling, ICIDS 2014, Singapore, Singapore, November 3–6, 2014. Proceedings. Cham: Springer.
- Pujol-Tost, Laia, Maria Roussou, Olivier Balet, and Stavroula Poulou. 2013. "Personalizing Interactive Digital Storytelling in Archaeological Museums: The CHESS Project". In Archaeology in the Digital Era. Volume II: e-Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology, Southampton, 26–29 March 2012. Graeme Earl, Tim Sly, Angeliki Chrysanthi, Patricia Murrieta-Flores, Constantinos Papadopoulos, Iza Romanowska, and David Wheatley, eds. Amsterdam: Amsterdam University Press, 77–90.
- Roberts, Catherine and Philip R. Stone. 2014. "Dark Tourism and Dark Heritage: Emergent Themes, Issues and Consequences". In Displaced Heritage: Dealing with Disaster and Suffering. Ian Convery, Gerard Corsane, and Peter Davis, eds. London: Boydell and Brewer, 9–20.
- Rose, Julia. 2016. Interpreting Difficult History at Museums and Historic Sites. Maryland: Rowman and Littlefield Publishers.
- Roussou, Maria. 2001. "The Interplay between Form, Story, and History: The Use of Narrative in Cultural and Educational Virtual Reality". In Virtual

Storytelling. Using Virtual Reality Technologies for Storytelling. International Conference ICVS 2001. Avignon, France, September 27–28, 2001. Proceedings. Olivier Balet, Gérard Subsol, and Patrice Torguet, eds. Heidelberg: Springer Berlin, 181–190.

- Runnel, Pille and Pille Pruulmann-Vengerfeldt. 2012. "Theorising Museum Participation". In The Transformative Museum: Proceedings of the DREAM Conference. 23–25 May 2012, Roskilde University, Denmark. Erik Kristiansen, ed. Odense: DREAM, 325–334.
- Sandell, Richard. 2007. Museums, Prejudice and the Reframing of Difference. London: Routledge.
- Screven, Chandler G. 1992. "Motivating Visitors to Read Labels". ILVS Review 2 (2): 183–211.
- Serrell, Beverly. 1996. Exhibit Labels: An Interpretive Approach. California: AltaMira Press.
- Silvén, Eva and Anders Björklund, eds. 2006. Difficult Matters: Objects and Narratives that Disturb and Affect. Stockholm: Nordiska Museet.
- Simon, Nina. 2010. The Participatory Museum. California: Museum 2.0.
- Springer, Julie, Sara Kajder, and Julia Borst Brazas. 2004. "Digital Storytelling at the National Gallery of Art". In *Museums and the Web* 2004. Selected Papers from an International Conference. David Bearman and Jennifer Trant, eds. Arlington: Archives and Museums Informatics, 123–131.
- Stone, Philip and Richard Sharpley. 2008. "Consuming Dark Tourism: A Thanatological Perspective". Annals of Tourism Research 35 (2): 574–595.
- Tarlow, Peter E. 2005. "Dark Tourism: The Appealing 'Dark' Side of Tourism and More". In Niche Tourism: Contemporary Issues, Trends and Cases. Marina Novelli, ed. Oxford: Elsevier, 47–57.
- Tinning, Katrine. 2017. The Ambivalent Potentiality of Vulnerability: Museum Pedagogy in Exhibitions on Difficult Matters and its Ethical Implications. Doctoral Thesis. Lund University. Available at: https://lup.lub.lu.se/ search/ws/files/34452788/The\_Ambivalent\_Potentiality\_of\_Vulnerability.\_Katrine\_TInning.pdf (accessed 21 April 2023).
- Tøndborg, Britta. 2013. "The Dangerous Museum. Participatory Practices and Controversy in Museums Today". Nordisk Museologi 2: 3–16.
- Violi, Patrizia. 2012. "Trauma Site Museums and Politics of Memory: Tuol Sleng, Villa Grimaldi and the Bologna Ustica Museum". *Theory, Culture and* Society 29 (1): 36–75.
- Williams, Paul. 2007. Memorial Museums: The Global Rush to Commemorate Atrocities. Oxford: Berg Publishers.
- Witcomb, Andrea. 2013. "Understanding the Role of Affect in Producing Critical Pedagogy for History Museums". *Museum Management and Curatorship* 28 (3): 255–271.
- Wong, Amelia. 2015. "The Whole Story, and Then Some: 'Digital Storytelling' in Evolving Museum Practice". In MW2015: Museums and the Web 2015. Chicago. Available at: https://mw2015.museumsandtheweb.com/paper/ the-whole-story-and-then-some-digital-storytelling-in-evolving-museum-practice/ (accessed 21 April 2023).

#### APPENDIX

# A: Category headings, on-site descriptions and number of Pasta items in alphabetical order (2019)

Category headings	
Animals: Far from your cute, cuddle-worthy pets, these pastas involve animals.	
Beings: Pastas that involve some unknown being, often with malicious intents.	2,785
Books: Pastas that involve some manner of books that isn't a diary or a journal.	97
Christmas: For when the world is silent, the fire crackles, and the snow is splattered in red.	41
Computers and Internet: Pastas with elements of computer weirdness, like cursed files, programs or websites.	507
Cryptids: A collection of unsettling cryptozoological-related stories.	122
Demon/Devil: Vile entities from the abyssal depths.	678
Diary/Journal: Pastas with a personal twist, meaning they involve someone's diary or journal.	694
Disappearances: People vanish every day. Perhaps the answers to some of their whereabouts can be found here.	811
Dismemberment: Sometimes, in pastas, people just go to pieces.	1,155
Dreams/Sleep: For those pastas that involve getting a little shut-eye then possibly never waking up.	1,112
Ghosts: How can you have a Creepypasta site without ghost stories?	968
Gods: Stories involving gods and deities.	185
Halloween: For the most horrifying time of the year, Halloween; when evil powers are at their strongest, and the boundaries between the living and dead are at their weakest.	68
History: Pastas about historic events.	351
Holders: They must never come together, for these are the original Holders series stories.	281
Hospitals: Pastas about hospitals.	76
Items/Objects: A breed of pasta that is centred around an item which is haunted, has a dark past or causes horrible things to happen.	872
Lost Episodes: These stories typically are geared towards revealing information about missing, lost, or otherwise previously unheard-of episodes, movies, or other digital media through a special screening or meeting usually directed to a very small group of people, and even one person. Note: we no longer accept Lost Episode pastas.	46
Lovecraftian: IA! IA! This kind of pasta fhtagn for you to read them.	179
Memes: Pastas that are well-cemented in 4chanian lore or are a certified meme.	100

Mental Illness: Pastas involving various levels of insanity.	
Military: Like the ramblings of an insane vet, these stories are interesting, scary, and about the military in some way.	236
Mirrors: These kinds of stories typically involve mirrors and their strange abilities.	192
Monsters: Things that go bump in the night but are neither beings nor demons/ devils.	989
Music: 🖍 For stories about music.	171
Nature: Pastas set or about the natural world, tapping into your primal fears.	333
NSFW: Pastas that might get you in trouble with your boss at work. Proceed with caution.	269
Photography: Pastas pertaining particularly to peculiar photos.	110
Places: These pastas, usually told in the first-person perspective, delve into an out- of-the-ordinary experience someone had at a certain place, usually local. These pastas can also recollect the scary history that this place has had.	1,483
Poetry: These stories are the bowtie pasta to the atypical spaghetti of prose pastas. These stories are poems.	269
Pokémon: Pastas involving the many multitudes of all 646 'mons. May Arceus have mercy on your soul. Note: we no longer accept Pokémon pastas.	98
Reality: Creepypastas made to destroy a pretence about reality you might have had. This may be through telling you that something you thought never existed or happened on a regular basis, in fact, does.	1,163
Ritual: All those pastas involving rituals; don't blame us should anything go wrong.	442
Science: Everything from insane experiments to scientifically based mind-fuckery is located here.	465
Space: Pastas taking place or dealing with the final frontier, outer space.	149
Television: If you're looking for TV weirdness, this is your category. If a pasta is here, it's not in the "Lost Episode" category.	142
Theory: Stories that look at something from a different perspective.	282
Troll Pasta: These are less horrible versions of Horrible Troll Pastas.	21
Vehicles: These stories are either based around cars, trucks, etc. or take place primarily inside them. Examples in popular literature would be the novel Christine or the movie Maximum Overdrive.	117
Video Games: Generally, these are stories about strange goings-on in video games.	148
Weird: Stories that can't quite be categorised any other way.	
Zelda: Open this chest and you might find an actual piece of a heart. Scary stories involving the Legend of Zelda franchise. Sorry, we no longer accept Zelda pastas.	

#### B: Tools used for the data analysis

- Anthony, Laurence. 2019. AntConc (Version 3.5). Tokyo, Japan: Waseda University. Available at: https://www.laurenceanthony.net/software (accessed 21 April 2023).
- Demšar, Janez, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. 2013. "Orange: Data Mining Toolbox in Python". Journal of Machine Learning Research 14 (August): 2349–2353.
- Sinclair, Stéfan and Geoffrey Rockwell. 2016. Voyant Tools. Available at: https://voyant-tools.org/ (accessed 21 April 2023).

### THE GENDER OF AI-GENERATED ARTS IN SARAH KENDERDINE'S "DEEP FAKES" AND AISHATU GWADABE'S AI ARTIVISM

#### NATASA THOUDAM

School of Liberal Arts and Digital Humanities-Interdisciplinary Research Platform, Indian Institute of Technology Jodhpur Jodhpur, Rajasthan, India nthoudam@iitj.ac.in

This paper compares the AI-generated art practices of Sarah Kenderdine and Aishatu Gwadabe. Kenderdine redefines the term "deepfake" to arrive at her new media art practice on digital replicas for which she uses the eponymous term "deep fake". In contrast to Kenderdine's digital replicas is Gwadabe's Afrofuturistic AI artivism. In Gender and Technology: A Reader, Nina E. Lerman, Ruth Oldenziel, and Arwen P. Mohun (2003) argue how technology is gendered by invoking as well as questioning the two terms "masculine technologies" (37) and "feminine technologies" (13). They further critique the translation of this gendered practice of categorising technologies into a racial one. Building upon their arguments, Jennifer Terry and Melodie Calvert (1997) in Processed Lives: Gender and Technology in Everyday Life, in fact, conflate the distinction between gender and technology by stating that "gender" is "itself a technology" - "a technology which produces, among other things, men and women" (4). They further urge for the need to reconceptualise "gender in cyberculture" and "the possibility of moving beyond the binary opposition between masculinity and femininity" (5-6). In light of these arguments, this paper re-examines the gendered relationship between women and technology in the production of the AI-generated arts of Kenderdine and Gwadabe.

Keywords: Afrofuturism, AI artivism, Aishatu Gwadabe, deep fakes, Sarah Kenderdine

#### INTRODUCTION

This paper examines the gendered relationship between women and technology in the production of AI-generated art by two women practitioners. The first practitioner is Sarah Kenderdine, who is a digital museologist and curator born in New Zealand and based in Switzerland. The other is Aishatu Gwadabe (aka Aishatu Ado), who is a peace technologist, AI artist and storyteller born and based in Germany but traces her origins to Nigeria. The paper also explores their disparate methods. While Kenderdine's method redefines the term "deepfake" (one word) to arrive at her new media art practice (AI art practice) on digital replicas for which she uses the eponymous term "deep fake" (two words), Gwadabe practices "AI artivism" (art activism that makes use of AI) with her born-digital artwork to expose racial biases of AI. The paper attends to two questions through these two practitioners. One is whether AI and its art practised by these two women are gendered. The other question is, if so, how are they gendered? To attend to these questions, the paper begins with an examination of the problematic relationship between gender and technology highlighted in Gender and Technology: A Reader by Nina E. Lerman, Ruth Oldenziel, and Arwen P. Mohun (2003) and the blurring of the line separating the two by Jennifer Terry and Melodie Calvert (1997) in Processed Lives: Gender and Technology in Everyday Life. From these general observations but within the specific context of North America, the paper attempts to trace a history of the emergence of AI-generated art, which, as argued, is also gendered. It, then, interrogates the artistic methods of these practitioners through the lens of gender.

#### **GENDER AND AI ARTS**

In *Gender and Technology*: A *Reader*, Nina E. Lerman, Ruth Oldenziel, and Arwen P. Mohun (2003) explore the intricate relationship between gender and technology. In the introduction of this book, through a critical reading of a song lyrics titled "Mister Rogers' Neighborhood",<sup>17</sup>

<sup>&</sup>lt;sup>17</sup> I'd like to be just like my Dad,

He's handsome and he's keen;

they argue that "the material world we inhabit and the gender roles we teach our children" and "childhood activities" reinforce gendered stereotyping as well as dichotomisation (1-2). Highlighting the process of gendering of technologies, these song lyrics make them raise these pertinent questions: "Are cars masculine technologies?" or "Are beds and stoves technologies?" If so, "are they [beds and stoves] feminine technologies"? The key terms here are "masculine technologies" and "feminine technologies". Lerman, Oldenziel, and Mohun seem to agree with the sociological distinction between the two; that is, "gender as a way of analyzing the social side of human activity and technology as a component of the physical world around us" (2). Yet, they still contend that "gender and technology are closely related" (2). To elaborate on their argument, they also cite historical instances in America, where "new digital technologies were named after the workers" (6). For example, women who "controlled" typewriters were referred to as "typewriters" (6); those who operated "the early telephone operating 'switches'" were also referred to as "switches"; and "(female) mathematicians whose work enabled missiles to launch accurately were called 'computers" (6). This resonates with the project on female computers by Lily Hibberd, Toner Stevenson, and Din Heagney. The project traces the story of "Sydney Female Computers" (Female Computers. Les Calculatrices).<sup>18</sup> These women "worked in over 20 observatories around the world from the late 19th century to create a catalogue and map of 6.5 million stars". Moreover, they were "[t]hought of as mere calculators" and "their labour was largely overlooked and eventually forgotten when the endeavour was finally abandoned" (ibid.).

He knows just how to drive the car,

And buy the gasoline.

I'd like to be just like my Mom,

She's pretty and she's nice;

She knows just how to make the bed,

And cook things out of rice

<sup>(</sup>Lerman, Oldenziel, and Mohun 2003: 2).

<sup>&</sup>lt;sup>18</sup> https://www.lilyhibberd.com/Female\_computers-menu.html.



Figure 1: "The Universe into a Cartesian Grid": How the 'Female Computers' Measured the Stars. Source: Screenshot of an image enlarged from the site: https://lilyhibberd.com/Female\_computers-1.html. Note: The original source is "'A Section of the Constellation Cygnus', 1885, by Paul Henry Collection Metropolitan Museum. Public domain" (https://lilyhibberd.com/Female\_computers-1.html).

In the words of Joan Rothschild who reviewed *Gender and Technology*: A *Reader*, the book "focuses on the reciprocal relationship between gender and technology in the specific historic and cultural context of North America from 1850 to 1950" (2004: 908). She further adds that "several articles illustrate the ways gender analysis not only questions gender boundaries themselves but also interrogates those of race, class, and ethnicity as these identities and power configurations intersect when humans engage in making, using, and shaping technologies" (ibid.). Hence, if this book critiques the categorisation of technologies as "masculine" and "feminine" and technologies being "named after the workers"; then, it also dismisses the racial practice in the Southern United States where "public restrooms were marked 'white' and 'colored" (Lerman, Oldenziel, and Mohun 2003: 6).

Another book that builds upon these arguments is Jennifer Terry and Melodie Calvert's Processed Lives: Gender and Technology in Everyday Life (1997). Also theorised in the context of North America, this book asks another set of related questions: "Are technologies themselves gendered? By what assumptions can we, for example, entertain the claim that guns are masculine machines and curling irons are feminine machines? Do classifications of this sort depend on the design of technology, or on its users?" (5). Blurring the distinction between the two categories of gender and technology and informed by Donna Haraway's insistence, in their words, that "feminists cannot afford to stay out of the game of technoscience", they insisted on moving beyond the "essential properties" of "thinking of things" and to "think of gender in cyberculture" that enabled "the possibility of moving beyond the binary opposition between masculinity and femininity" (5–6).

From these observations on technology in general, this paper moves on to trace the history of the emergence of a particular technology, namely, AI and its role in artistic production. The section titled "Portrait of the Computer as an Artist" in Arthur I. Miller's Artist in the Machine: The World of AI-Powered Creativity (2019) somewhat attempts to chronicle this history. For Miller, the history of AI art began in the 1960s with A. Michael Noll's "computer art" called *Gaussian Quadratic*. Miller describes the processes involved in this "creation" thus: "His [Noll's] program produced dots distributed in a way prescribed by a mathematical equation called a Gaussian distribution, which generates bell-like curves. He then connected the dots from bottom to top in a random way, producing a continuous zigzag".

The next event that Miller discussed happened in 2015 when Alexander Mordvintsev of Google attempted to create dream-like images using *Convolutional Neural Networks* (ConvNets).<sup>19</sup> Mordvintsev created a "nightmare beast" with what was later called *DeepDream*, a nomenclature for this new "machine" that could create art, resulting

<sup>&</sup>lt;sup>19</sup> Miller explains this term by drawing an analogy between neural networks and neurons of the human brain: "Neural networks are designed to mimic the brain. (...) Artificial neural networks are designed to replicate the activities of the brain and to cast light on how the brain works." He, then, compares convolutional neural networks with neurons of the human eye: "Convolutional neural networks are a specialised form devoted mainly to vision, able to recognize objects and spot patterns in data. The neurons are arranged in a similar way as in the eye" (2019).

from his collaboration with another Google engineer Mike Tyka. After this, Miller documents how Tyka went on to form Artists + Machine Intelligence (AMI) in Google with Blaise Agüera v Arcas. The artistscholar Memo Akten, influenced by DeepDream, exhibited Learning to See: Hello, World! in 2017. This piece, according to Miller, demonstrates "how neural networks come alive as they begin to observe the world around them through the data they are fed". Another instance of "how a neural net awakens" discussed by Miller is "Damien Henry's computer-generated Music for 18 Musicians - Steve Reich". Another machine learning technology that simultaneously emerged along with DeepDream was the Generative Adversarial Network (GAN),<sup>20</sup> which was invented by Ian Goodfellow of the University of Montreal in 2014. Miller cites the example of Tyka's Portraits of Imaginary People as a GAN-generated art. A further development of GANs, observed by Miller, is Phillip Isola's CGANs (Conditional Generative Adversarial Networks).21 This system is named as Pix2Pix by Isola. Here, in this gendered history of "machine creativity" or "computational creativity" by Miller, one of the female artists discussed is Anna Ridler who made use of Pix2Pix to create "an abstract film" - a piece "exploring the 1928 silent film Fall of the House of Usher, adapted from Edgar Allan Poe's short story, using her own images".

One of the derivatives of GAN that Miller lists is Jun-Yan Zhu's Pix2Pix CycleGAN. Jun-Yan's quest was "to find an image of a horse

<sup>&</sup>lt;sup>20</sup> According to Miller: "GANs are a way of training the computer to create realistic images entirely by itself" (2019). He further states that "GANs are based on game theory" and explain their underlying mechanism wherein "[t]here are two dueling networks, the discriminator (D) and the generator (G)", both of which "are deep neural networks". The difference between them is that "D is fed with images of the real world, from a dataset such as ImageNet", while "G begins to generate images out of a first layer that is latent space, made up of noise, like randomly situated dots". In this arrangement, "G is like an art forger, while D operates like a detective evaluating the images generated by G", deciding "whether the image it receives from G is realistic, based on the images it has been fed, or whether it is unrealistic – that is, fake". Interacting with D, G "starts to learn how to produce images that look real". In this interaction between D trying "to distinguish realistic from unrealistic images" and G trying "to create images that will fool the discriminator into thinking they are real", Miller observes that "a stable situation emerges in which G's images actually look like the ones in D's training set". He named this situation after the term "the Nash equilibrium" from game theory.

<sup>&</sup>lt;sup>21</sup> Miller distinguishes them from GANs by explaining that "[t]hey are conditional because instead of starting the generator network (G) from noise, from nothing, they condition it by using an actual image" (2019).

that exactly matched an image of a zebra, in exactly the same position". To achieve this objective, he ran "the cycle over and over again, using algorithms to reduce the loss of consistency in both directions". Here, "[h]e tried inserting an image of a horse into several unpaired images of zebras". Ultimately, "[a]fter a number of cycles", he achieved his objective. The other is Ahmed Elgammal's CANs (*Creative Adversarial Networks*), which is similar to GANs in having two networks: D and G. The distinguishing feature of CANs from GANs according to Miller was that "D is trained on the WikiArt dataset and learns to discriminate between art and nonart", while "G is untrained, in a state of pure noise, in latent space".

The year 2018, for Miller, marks the entry of GANs into the art market with the auction of a GAN artwork, Portrait of Edmond de Belamy, created by the all-male French collective Obvious. The invention of Simon Colton's The Painting Fool could be considered a landmark in this history of computer creativity, where there is a blurring of the separation between a human artist and a digital artist (here The Painting Fool). Miller invokes the image of a "struggling" software (as a digital artist) to highlight its similarity with a human artist. He argues that: "Although the program was created by Colton, The Painting Fool's artwork is its own". To elaborate on his contention that the artwork created should be attributed to the programme/software/digital artist The Painting Fool, Colton associates this with the larger objective of his project, which is his interest "in the simulation of particular behaviours associated with creativity, such as imagination, appreciation and intentionality". Moving beyond this world-recognised digital artist, Miller also brings in robot artists such as Hod Lipson's "self-aware" robots and Patrick Tresset's "sketching robots, collectively called Paul" (ibid.). Finally, since 2016, there have been at least three Annual Robot Art Competitions and Exhibitions.<sup>22</sup> Excepting the tokenish reference to Ridler, Miller's survey highlights a gendered history of AI-generated arts. The glaring absence of women practitioners in this survey is shown as a historical fact as both the human artist and the digital artist invoked are imagined as European as well as male. It is within this gendered history of absence that this paper discusses the two women practitioners.

<sup>&</sup>lt;sup>22</sup> https://robotart.org/.

# GENDER IN KENDERDINE'S "DEEP FAKES" AND GWADABE'S AI ARTIVISM

One of the texts used for analysis in this paper is a YouTube video recording of Sarah Kenderdine's presentation "Deep Fakes: Art and Its Double – An Exhibition" from a panel titled *Preservation of Cultural Material*. This text is then compared with Aishatu Gwadabe's YouTube video recording of a keynote lecture she addressed for AIxDesign titled "AI Artivism: For Peaceful, Just, and Pluriversal World".

As indicated in the caption of Kenderdine's presentation video on YouTube, her exhibition "poses crucial questions about the potency of digital replicas to absorb audiences in enduring emotional encounters with universal art treasures" (Deep Fakes: Art and Its Double – The Exhibition (Sarah Kenderdine) 2022). Kenderdine clarifies that even though the exhibition has taken the "provocative title: deep fake" (two words), "it opposes the popular usage of deepfakes [one word] for manipulations and misinformation"; rather for her, it explores "very different perspectives: [that is] reimagining objects through advanced computational techniques".

As she provides her meaning of "deep fake" (two words again), her presentation in the background highlights a quote from the American art historian Rosalind Krauss which reads: "The image of the counterfeit functions as a complex figure, a kind of pretzel in which true and false chase each other's tails. Glass and gold, pure and impure, begin to reflect on one another, as in a hall of mirrors". A series of photos of her exhibition constitutes her presentation in the video. In this presentation, she explores six installations from a total of twenty-one in this exhibition. Putting forward a case for digital replicas, her concept of "deep fakes" allows her to reimagine "objects" as well as "objecthood remediated through participatory interphases such as mixed, augmented, and virtual reality". She adds that the exhibition is also "grappling with the digital materialities that objects possess in post-original form". In other words, what she examines are "objects": artworks as "objects" while also emphasising a particular temporal and critical stage of these objects that is "post-original form". Through the notion of "augmented replica", she contends, "machine learning antagonizes outdated notions of authority, authenticity, and access". She also highlights the decolonising feature of these digital facsimiles as "they defy hegemonic narratives helping to liberate things from their colonial entrapments confronting authoritative discourses, historical sedimentation, and contested social relations". This leads her to arrive at a conclusive statement that "digital copies have enabled communities to become more resilient to loss" as "they can also provide reservoirs of cultural memory and instruments for those on the margins to speak back to their oppressors". These statements were made concerning her installation titled *Speaking Back* on the Black Lives Matter campaign.

Arguing further on her case for digital copies, she identifies one vital question facing "deep fakes" which is "how the inscription of digital counterparts instead of the originals is exploding systems and codes of ownership, custodianship, and repatriation". To support this argument which was made citing her installation titled Pure Land Augmented Reality Edition, she states that "immersion [immersive experiences enabled by these digital counterparts] is capable of producing a sensation of presence, that is, the sensation of being there through which participants engage in as actors with synthetically generated world". Furthermore, for her, "high fidelity digital fac similes no longer merely act as complements to real objects but have profound effecting presence in their own right". One such example she cites is of the installation titled The Next Rembrandt by the creative agency Wanderman Thompson that shattered "the inviolability of the original", while another installation titled 89 Seconds Atomized by Eve Sussmen employs blockchain to demonstrate "collective ownership". Finally, the last installation she discussed was titled Helin by Christian Mio Loclair. Carved by a robotic arm, it is an "exempla [sic] of first-generation AI" that "challenges the status of art in the absence of a human author;" thereby, "probing individual human and artistic intent" (ibid.).

Through this description of her exhibition on "deep fakes" by Kenderdine herself, what emerges is Kenderdine as a collector or assembler of digital artworks or replicas or facsimiles. In contrast to this collector of copies, Gwadabe is a creator of "original" AI artwork. In her keynote lecture, she describes herself as an "AI artivist", which entails fusing three roles of a "peace technologist", an "AI visual artist" and a "critical AI ethicist" (AIxDesign Keynote – AI Artivism with Aishatu Gwadabe 2021). As an AI artist, she has "been exploring machine learning techniques" in order to explore issues of knowledge and how she can "explain theoretical concepts with an image". Through the influence of her journalist father, she learnt to see "the effectiveness of storytelling". Another influence on her art practice and its politics is "the idea of Afrofuturism or Black science fiction", which she invoked to critique the genre of science fiction. Her aversion to science fiction stems from her observation that "people of colour are not found in the imagination of the future" as their stories "did not include me". This led her to Afrofuturism which allowed "alternative futures". She, thus, defines AI artivism as "doing activism through artificial intelligence and art". However, she does not fall into the trap of techno-positivism here as she is aware and also wishes to raise awareness about AI biases and the need for an "inclusive" and "just" AI.

These biases take the form of either racial discrimination through her example of a soap dispenser that failed to recognise a dark-skinned hand, or they are gendered as in the historical case of female human computers. In the lecture, she puts her art practice within the domain of design, and in light of this intersectional discrimination, she advocates the need to reimagine this domain in a postcolonial context. This reimagination for her includes a shift from the Descartian "ego-logical" (human-centric) to an "eco-logical" (environment-centric) world. For her, the "ego-logical" world is "authoritarian, dualistic, unsustainable, delusional, mechanistic, self-destructive, unwise, imbalanced, and power-seeking", while the "eco-logical" world is characterised by these words: "democratic, holistic, sustainable, compassionate, natural, regenerative, wise, balanced, and interdependent". In other words, she sees the "ego-logical" world as "hierarchical" with "the human [also man] on top" while "the female figure" along with all other living beings are below him. Hence, the "ego-logical" world is also gendered for her. This world, according to her, reflects René Descartes' "Cogito, ergo sum" dictum (I think, therefore I am) as well as the psychological idea of "ego" as the "I", which lies at the centre. "Eco-logical", in contrast, is "non-hierarchical", where "every being is the same" with equal importance. It is also a "more environment-centred" approach. This is also reflected in her idea of "making kin with the machine".<sup>23</sup> Continuing with her support for Afrofuturism, she posits the Ubuntu philosophy of "we think, therefore we are" in opposition to the Descartian dictum. She argues that this philosophy with peace as her compass, design and interest enables her to achieve "pluriversality" which is pitted against universality and imagines the possibility of "having many worlds in one world". She demonstrates this philosophy through two of her AI-generated artworks in this presentation: "Fulani Woman" and "African Epistemology" (ibid.).

To examine these two art practices through the lens of gender, the paper now looks for the subjects of Kenderdine's immersive "deep fakes" and Gwadabe's Afrofuturist AI artivism. Since data is an important component of machine learning techniques used for the AI artworks created by Gwadabe and those curated by Kenderdine, the paper discusses the problematic characteristics of a data subject. In the context of data protection rights under the European General Data Protection Regulation (GDPR), the issue noted by political theory scholar Aisha Kadiri is the emergence of "a [White] liberal [individual] data subject" that "digitally" enables racism and sexism (2021: 6). To attend to this problem, she finds an alternative imagination of this subject through Afrofuturism, even though there is criticism of its "focus on the Western lens and its decentering of African imaginaries" (Okorafor and Wabuke cited in Kadiri 2021: 7 [Fn. 3]). The resultant data subject is "a liberated and intersectional [collective] data subject" (Kadiri 2021: 8). For Kadiri, the term intersectional is derived from Crenshaw's "intersectionality", that is "the interaction of different demographic factors that work together to either empower or oppress" (9). She finds this subject as enabling because s/he "accounts for interconnectedness, unsettles predefined categories, and acknowledges the structural aspect of discrimination" (9). This subject also counters "the historical exclusion of Black people from the category of subjects" and the underlying "notion of denied subjectivity" observed from Toni Morrison's statement that "Black people" are "objects of history, not subjects within it" (cited in Kadiri 2021: 11).

<sup>&</sup>lt;sup>23</sup> The term "making kin" was introduced by anthropologist and feminist scholar Donna Haraway, with further insights available in her influential work Staying with the Trouble: Making Kin in the Chthulucene (2016).

NATASA THOUDAM

In terms of the subjects imagined within the problematics of this data subject, an obvious distinction can be seen between Kenderdine's liberal individual data subject and Gwadabe's Afrofuturist intersectional collective data subject. However, the dynamics between these two subjects are not as straightforward as Kadiri has highlighted. Kenderdine's data subject is a replica, a reproduction and a copy, while Gwadabe's is the born-digital original. This is a clear inversion of the hierarchy that exists between a liberal subject and a "Black object" brought about by digitally enabled racism. In fact, Kenderdine's focus is not on the creator subject but on the created object. In contrast, Gwadabe invokes both creators and the created alike. Gwadabe's AI artivism against data biases in the production of AI technologies finds no mention in the replicative technological processes of Kenderdine's "deep fakes" even when the latter is in favour of "an inclusive difficult heritage" that is characterised as "dissonant", "negative" and "undesirable"; that "reveals unheard or silenced stories"; "challenge preconceptions"; "encourage emotional and empathetic responses"; and "create a sense of presence, immersion or embodiment, and ultimately provide unique experiences" (Stylianou-Lambert, Bounia, and Heraclidou 2022: 1-2).

This also refers to a distinction between their two methods. Gwadabe's art practice is still within the domain of "meaning culture", a term borrowed from literary theorist Hans Ulrich Gumbrecht, wherein "knowledge is produced by a subject who is observing, rather than participating, in the world" (cited in Stylianou-Lambert, Bounia, and Heraclidou 2022: 13). This is evident in her invocation of Descartes, Ubuntu philosophy and African epistemology, and the need to "explain theoretical concept with an image" as the driving force for her visual AI art. Gumbrecht also argues for the need to shift from "a 'meaning culture' to a 'presence culture'" (cited in Stylianou-Lambert, Bounia, and Heraclidou 2022: 13). Kenderdine's immersive "deep fakes" demonstrate this "presence culture" whose "main dimension" is "space and more particularly the relationship between human bodies and the world around them" (Stylianou-Lambert, Bounia, and Heraclidou 2022: 13). Accordingly, "emerging, and especially immersive, technologies [such as Kenderdine's "deep fakes"] have the potential to provide experiences that are not exclusively related to meaning, cognitive knowledge and
interpretation but can also create a sense of presence, immersion, and embodiment" (ibid.).

From this pondering on the data subject, the paper asks a crucial question regarding gender. The category gender is used as a lens to examine the AI art practices of two women subjects. However, it is not the same as a feminist lens. While this category critiques the gendering of technology, it does not reinforce a gender binary as it further questions institutions such as patriarchy, racism, sexism, classicism, colonialism, and liberalism, being aware of the fact that both genders are equally implicated within these institutions. Both Kenderdine and Gwadabe are attentive to the gendered nature of their contexts, namely digital heritage and AI technology. Solutions they proffer are "an inclusive difficult heritage" for Kenderdine and the philosophy of Afrofuturism for Gwadabe. Both are advocates of intersectionality. Aligning with Gwadabe, Kenderdine also critiques racism through her installation on Black Lives Matter. Still, her critique is within a "White" European epistemological framework. Similar criticism has also been levelled against Afrofuturism.<sup>24</sup> Moreover, the statuses of the reproduced digital double and the "Black" Afrofuturist data subject are always in an oppositional relation to the original and the "White" liberal data subject, respectively. The question that remains is whether the intersectionality of the two artists entails transcending the binaries of men-women, masculinefeminine, original-replica, production-consumption, Western-African, White-Black, and subject-object or reinforcement of them while the digital space could be a site for their dissolution.

#### **REFERENCES LIST**

- "AIxDesign Keynote AI Artivism with Aishatu Gwadabe". 2021. Available at: https://www.youtube.com/watch?v=e3NACCtzOGc (accessed 11 October 2022).
- "Deep Fakes: Art and Its Double The Exhibition (Sarah Kenderdine)". 2022. Available at: https://www.youtube.com/watch?v=Ae5tLCzpPw4 (accessed 11 October 2022).

<sup>&</sup>lt;sup>24</sup> On the criticism and debate around Afrofuturism, see Martin Bernal's Black Athena: The Afroasiatic Roots of Classical Civilization and Mary Lefkowitz's Not Out of Africa: How Afrocentrism Became an Excuse to Teach Myth as History.

- "Female Computers. Les Calculatrices". Available at: https://www.lilyhibberd.com/Female\_computers-menu.html (accessed 17 December 2022).
- "Introduction: Emerging Technologies, Museums and Difficult Heritage". 2022. In Emerging Technologies and Museums: Mediating Difficult Heritage. Theopisti Stylianou-Lambert, Alexandra Bounia, and Antigone Heraclidou, eds. New York, NY: Berghahn Books, 1–19.
- "Introduction: Interrogating Boundaries". 2003. In *Gender and Technology:* A *Reader*. Nina E. Lerman, Ruth Oldenziel, and Arwen P. Mohun, eds. Baltimore, MD and London: The Johns Hopkins University Press, 1–11.
- "Introduction: Machine/Lives". 1997. In Processed Lives: Gender and Technology in Everyday Life. Jennifer Terry and Melodie Calvert, eds. London and New York, NY: Routledge, 1–13.
- Kadiri, Aisha P.L. 2021. "Data and Afrofuturism: An Emancipated Subject?" Internet Policy Review 10 (4): 1–26.
- Miller, Arthur I. 2019. "Portrait of the Computer as an Artist". In Artist in the Machine: The World of AI-Powered Creativity. Cambridge, MA and London: The MIT Press. E-book.
- Rothschild, Joan. 2004. "Review of Gender and Technology: A Reader". Technology and Culture 45 (4): 908–910.

### LIMITATIONS

The paper aims to explain why there has been no discussion on new AI technologies such as AI art generators. It justifies this absence with the fact that it is beyond its scope as both Kenderdine and Gwadabe used machine-learning (or deep-learning) techniques for the artworks they respectively curated and created. Hence, the history that was traced in this paper focused on this particular kind of art generation. Still, the paper is aware of the emergence of AI art generators and demonstrates them through an authorial exercise (see Figures 2 and 3).



Figure 2: An Authorial Exercise: Deep Fake. Source: The image was created using the text prompt "Female Deep Fake" in the cyberpunk style on https://creator.nightcafe. studio/.



Figure 3: An Authorial Exercise: Afrofuturist AI. Source: The image was created using the text prompt "Afrofuturist AI" in the steampunk style on https://creator.nightcafe. studio/. This authorial exercise further complicates the already blurred binary between an original and its replica and the issues associated with it.

#### **APPENDIX**

### **AI AND CREATIVE ARTS**

A concept note written but never published for a specialisation offered by the Centre of Excellence on Art and Digital Immersion, School of Liberal Arts, Indian Institute of Jodhpur, India.

"CAN COMPUTERS BE CREATIVE?"

Arthur I. Miller raised this question in his book *The Artist in the Machine*: *The World of AI-powered Creativity*. It pertains to an ongoing debate between two groups of critics.

The group that favours computer creativity say the following:

"The brain is just a computer like any other (...) Traits previously considered innate to humans – imagination, creativity, even consciousness – may be just the equivalent of software programs." – Demis Hassabis

"Machines are already creative." - Ian Goodfellow, Google

"The more intelligent AI becomes, the more sophisticated its art will be." – Hod Lipson, Columbia University

"In principle, because the brain obeys the law of physics, computers can do anything the brain can do." – Murray Shanahan, Imperial College, London

"When we do art with machines, I don't think there is a very strict boundary between what is human and what is machine." – Blaise Aguera y Arcas, Google

The other group, against such creativity, claims:

"I'm going to be a hardliner and say that computers cannot be creative." – Allison Parrish, NYU

The timeline and milestones of computer creativity according to Miller are as follows:

1956: AI coined by John McCarthy

1965: First computer art:

(i) German: Frieder Nake and Georg Nees

(ii) American: A. Michael Noll

- 1968: The first computer program, AARON, that creates original artistic images
- 1970s: Earliest instance of computer music
- 1991: Computer-enhanced cello
- 1990s: Creative or generative algorithms for computer music
- 2003: François Pachet's *Continuator* (a system that can improvise with a musician)
- 2014: Ian Goodfellow invented generative adversarial networks (GANs)
- 2015: DeepDream by Alexander Mordvintsev Artists and Machine Intelligence (AMI) formed Mario Klingemann's Monalisa
- 2016: Exhibition: DeepDream: The Art of Neural Networks
  Angelo Semeraro's AI programme Recognition
  Leon Gatys' Neural Style Transfer algorithm
  Douglas Eck's Project Magenta (music composition by computer)
  Launch of WaveNet, NSynth, and Coconet
  Computer composing folk music
  Morwaread Mary Farbood's Hyperscore
- 2017: Arts Electronica exhibition in Linz showing Learning to See: Hello World by Memo Akten
  Damien Henry's computer-generated Music for 18 Musicians Steve Reich
  Mike Tyka's Portraits of Imaginary People
  Theresa Reimann-Dubbers' A(.I.) Messianic Window exhibited
  Jake Elwes' machine dream in Latent Space
  Phillip Isola's Pix2Pix
  Jun-Yan Zhu's cycleGAN
  Ahmed Elgammal's Creative Adversarial Networks (CAN)
  2018: Auction of Portrait of Edmond de Belamy created by a GAN
  - First digital artist: Simon Colton's The Painting Fool Artist robots: Hod Lipson and Patrick Tresset

Computer creativity in India:

- 2018: Art space hosting AI-related event: Nature Morte hosting Gradient Descent, curated by Harshit Agrawal Open Codes, an event of ZKM Centre for Art and Media, Karlsruhe, held in India since 2018
- 2019: Auria Kathi, an AI bot artist, developed by Fabin Rasheed and Sleeba Paul, given a public identity

Sahej Rahal's artwork Juggernaut made use of video game

2021: Raghava KK's La Petite Mort (an NFT with physical artwork) with Harshit Agrawal,

Abhijeet Satani, and Ben Tritt

Agarwal's participatory work (*author*)rise created in collaboration with Junichi

Yamaoka and Yasuaki Kakehi examine the concept of fragmented authorship of digital texts

Agarwal's Strange Genders uses GANs to raise questions about gender

Agarwal's first solo exhibition, EXO-stential – AI Musings on the Posthuman, hosted by Emami Art Gallery in Calcutta

Rahal's finalforest.exe is built on a video game

2022: Raghava's AI art project Cyborg Desires

Jenny Bhat "introduced augmented reality (AR) to her digital interactive art project 'MokshaShots'"

2023: Bhat's Metamind: Infinite Possibility "uses digital layers to bring motifs of infinity, metaverse, and the human mind to life", Jaideep Mehrotra's digital brush (active since 1990s) and the screenwriter Prateek Arora bring Sci-Fi and AI together

(Sources: https://mapacademy.io/article/ai-art-in-india/ and https://timesofindia. indiatimes.com/india/how-indian-artists-are-using-ai-ar-to-let-creativity-soar/ articleshowprint/98644302.cms.)

These timelines and milestones make us return to the question: "Can computers be creative?" This question is both the beginning and the end of the creative, ethical, sociological, and critical quests of this specialisation. Thus, this specialisation not only explores the relationship between AI and creative arts but also sees AI as a new emerging creative art.

## **TEACHERS AS CREATORS OF ETHNOGRAPHIC KNOWLEDGE –** PRESENTATION OF COLLECTED MATERIAL IN A DIGITAL ENVIRONMENT

### DANIJELA BIRT KATIĆ

University of Zadar Department of Ethnology and Anthropology Zadar, Croatia dbirt@unizd.hr

### MARTINA KRIVIĆ LEKIĆ

State Archives in Bjelovar Bjelovar, Croatia krivic.lekic@dabj.hr https://orcid.org/0000-0002-8826-5956

"Policies of Recording Ethnographic Materials from 1897 to 1954 – Teachers as Creators of Ethnographic Knowledge" (project acronym: UČeka) is an institutional project supported by the University of Zadar. The project primarily focuses on archival research of manuscripts written by teachers at the end of the 19<sup>th</sup> and the beginning of the 20<sup>th</sup> century. By reviewing, researching, and critically analysing the archival material containing ethnographic data the project aims to detect key factors that influenced aspects of the teacher's collecting and recording of ethnographic material in the period from 1897 to 1954. Along with evaluating the work of teachers in collecting and describing ethnographic data, our goal is to popularise and contribute to the understanding of the role ethnologists have in the contemporary context, as well as disseminate and popularise the topics of ethnology among other topics of research. At the beginning of the project, we decided that it was necessary to develop a way of displaying and sharing the collected archival material. Considering the diversity of the collected material and possibilities for the publication of digital collections and exhibits, the digital platform Omeka.net was chosen. The Omeka.net platform is a place where the materials of various institutions will be consolidated, presented and available for further interpretation and valuation.

Keywords: teachers, ethnography, digital platform, digital space, Omeka

# PRESENTING THE PROJECT TIMELINE, AIMS, ACTIVITIES, AND EXPECTED RESULTS

The undergraduate level of the Ethnology and Anthropology study programme at the University of Zadar consists of several courses where students are introduced to the development of disciplinary frameworks of Croatian ethnology. For the past ten years, Danijela Birt Katić, one of the co-authors of this paper, has been teaching a course called History of Croatian Ethnology. The goal of the course is to teach students about the developments of ethnological thought in Croatia with an emphasis on the contribution of Antun Radić. During the preparation for the lectures, reading various literature, but also conversations in which an effort was made to broaden the understanding of the scope of Radić's work, the topic of what would later turn out to be a scientific project was formed. From 2020 until 2021 Danijela Birt Katić was participating in the Erasmus+ project TRANSCA.25 The broad aim of that project was to think about how ethnology and cultural anthropology can contribute to the education of teachers and other educational workers with their knowledge. That combined formed an idea for the UČeka project with the focus on teachers as a specific group of collectors of folklore and ethnographic materials.

<sup>&</sup>lt;sup>25</sup> TRANSCA ("Translating Socio-Cultural Anthropology into Education") was an Erasmus+ project coordinated in Austria at the University of Vienna Department of Social and Cultural Anthropology, but also Department of Balkan, Slavic & Oriental Studies, University of Macedonia, Greece, Danish School of Education, Department of Educational Anthropology at Aarhus University, Denmark and Department of Ethnology and Anthropology University of Zadar, Croatia participated in the project. The main resource on this page is different concepts from sociocultural anthropology, which are prepared for teachers to use in their classrooms.

Since September 2021, this project has been carried out at the Department of Ethnology and Anthropology under the title "Policies of Recording Ethnographic Materials from 1897 to 1954. Teachers as Creators of Ethnographic Knowledge" (project acronym: UČeka). It is financed by the University of Zadar. The project's propositions called for experts from university departments to gather in the team, with great support from experts coming from cooperating institutions. Experience from previous projects guided the selection of team members, as well as their specific expertise and location. Members of the project team are PhD Marija Mičić Buterin, from the Department of Pedagogy, PhD Ines Cvitković Kalanjoš from the Department of Teacher Education Studies in Gospić, and Assistant at the Department of Ethnology and Anthropology Adriana Branka Pojatina. External collaborators engaged in the project are Martina Krivić Lekić, director of the State Archives in Bjelovar, PhD Karmen Turčinov, professor at the High School of Natural Sciences and Agriculture, Food and Veterinary in Zadar, and PhD Ivana Štokov, independent researcher with specific knowledge of GIS application.

As was previously emphasised, the overall aim of the project is to make research in the archival funds in order to determine whether there are manuscripts of ethnographic material written by teachers. Specifically, teachers who are in the process of collecting used Antun Radić's ethnographic methodological questionnaire, known under the title Osnova za sabiranje i proučavanje građe o narodnom životu.<sup>26</sup> The found material will serve in the interpretation and valorisation of the teacher's collecting work in the context of the policies of the time. These policies were shaped by guidelines developed through the initiatives of the Croatian Academy of Sciences and Arts, to which Antun Radić contributed greatly. The UČeka project covers a period of almost 60 years. We begin in 1897, when Radić's ethnographic method was first

<sup>&</sup>lt;sup>26</sup> The ethnographic methodological model for doing fieldwork was shaped by Antun Radić, better known as the brother of Stjepan Radić, one of the founders of the Croatian Peasant Party, but remembered in ethnological circles as the editor of the journal Zbornik za narodni život i običaje (Collection for Folk Life and Customs), published by the Croatian Academy of Sciences and Arts. The text was published in 1897 under the title Osnova za sabiranje i proučavanje građe o narodnom životu (Foundation for Collecting and Studying Materials on Folk Life), and to this day represents the most comprehensive and concise questionnaire for fieldwork in ethnology.

published, and end in 1954, when the collection policies that had been profiled up until that point changed, partly due to a change in journal *Zbornik za narodni život i običaje*, i.e. the change occurred in the editor's position. For the purpose of contextualising the interpretation and valorisation of the contribution of teachers, the manuscripts of teachers that are not stored in the Academy's archive or written exclusively according to the template set by *Osnova* were also taken into account. Therefore, school funds, and personal funds of individual people, stored in the State Archives in Bjelovar, the State Archives in Gospić, and the State Archives in Zadar were examined in detail, while in the Croatian State Archives, only specific teachers' funds will be examined.

We searched the archives for ethnographic memory books<sup>27</sup> with the assumption that we would find them only in the area that was on the territory of Banovina of Croatia (Banovina Hrvatska). The writing of the ethnographic memory books was prescribed by the decree of ban Ivan Šubašić in January 1940 as mandatory documentation in public schools. The goal of the writing was to acquaint the teacher with the everyday life of the peasants/people with the aim of accepting these people as social subjects, not objects that need to be completely re-educated. Immediately after the decree was adopted, the Department of Education published Instructions for Managing Ethnographic Memory Books. According to the instructions, teachers and directors of public schools were recommended getting to know the people "in order to prevent mistakes and disturbances that accompany the process of training and cultivation". It is pointed out that teachers are not sufficiently familiar with the life of children outside of school and incorrectly assess "some phenomena in children, they do not know the real cause". Given that it happened that some teachers, probably out of ignorance, inadvertently destroyed or participated in the destruction of traditional, indigenous folk peasant culture, collecting and writing down about the everyday life of the local village should have contrib-

<sup>&</sup>lt;sup>27</sup> The obligation to write school memory books was prescribed by the first school law "On the Organisation of Public Schools and Preparation for Public Teaching in the Kingdom of Croatia and Slavonia", in 1874. Dinko Župan writes that school memory books are an extremely important source for researching the history of education as well as the native or local history, among other things they contain otherwise scarcely available information about the history of the place, school life and the life of the local people (2010: 212).

uted to the multi-year efforts of the Peasant political party to "return" the folk culture to the people.

To write down the data, teachers were recommended to use Antun Radić's ethnographic methodological questionnaire as guidance. The project of writing ethnographic memory books was conceived over a four-year period, that is, the material collected during that period was supposed to serve in the planning of new educational programmes and policies. A famous ethnologist at that time, Milovan Gavazzi,28 was engaged in the promotion of the writing of ethnographic memory books. He edited several texts in teachers' journals, but individual teachers, such as Marijan Markovac, also contributed to the agenda (1940).<sup>29</sup> In August 1940, the ABC Club and the Ethnological Seminar organised the first Croatian ethnographic congress at the University of Zagreb, where there was a discussion on ethnographic memory books written by public school teachers. For this occasion, the ethnographic memory books were described as a pedagogical tool necessary for achieving the aims of school reform - and getting closer to the final aim of establishing a new, people's school. In the following years, the project of writing ethnographic memory books failed due to political changes caused by the beginning of the Second World War. Only a few of the memory books found stored in the archives point to the minimal maintenance of this practice.30

Elementary school funds examined in the State Archives in Bjelovar revealed a manuscript entitled *Village and Villagers* written by teacher Stjepan August about the village of Blagorodovac near Daruvar. The

<sup>&</sup>lt;sup>28</sup> Milovan Gavazzi is one of the most famous Croatian ethnologists of the 20<sup>th</sup> century. The range of his research is still an inexhaustible topic and a challenge for new generations of ethnologists who strive to understand, follow, and criticise his research approach.

<sup>&</sup>lt;sup>29</sup> The work of teacher and pedagogue Marijan Markovac is related to the first half of the 20<sup>th</sup> century. In addition to working as a teacher, for one period he was employed as a curator of the Croatian School Museum in Zagreb.

<sup>&</sup>lt;sup>30</sup> Field reports from that time write about the minimal engagement of teachers in writing the ethnographic memory books. A good example is the report of the school superintendent of the Gospić district, in the first half of the school year 1940/1941, only eleven out of thirty-four schools recorded work on the memory books. The data from the State Archives in Bjelovar also shows the poor engagement of teachers. Four ethnographic memory books have been preserved in a total of 248 primary school funds, with the fact that two have just been started, and the remaining two are not complete.

aspects of everyday life that Stjepan August recorded from today's perspective are valuable material that gives us an insight into local life a hundred or more years ago. The author recorded the making of the first settlement in the area, the immigration of the Czech and German population, the construction of the railroad, the clearing of oak forests, as well as customs throughout the year specific to the inhabitants of Blagorodovac. Stjepan August's text is the closest to the structure of school memory books, even though he does not mention any other writing model, not even Antun Radić. Despite this, the author provides rich material about the traditional culture of Blagorodovac. For these reasons, the manuscript was selected for publication as part of the activities of the UČeka project. After the process of transcribing the manuscript, followed by a detailed reading and comparison of the original with the transcription, the manuscript will be critically processed, prepared for publication, and printed, but also available in digital form on the omeka.net platform.<sup>31</sup>

### TEACHERS AS CREATORS OF ETHNOGRAPHIC KNOWLEDGE

In the chapter that follows we will closely take a look at the processes that influenced the creation of teachers as collectors of ethnographic materials in the beginning of the 20<sup>th</sup> century. Antun Radić's ideas and activities strongly marked the shaping of Croatian ethnology's research framework. In 1897, when the second issue of the journal *Zbornik za narodni život i običaje* was published, Radić became the editor-in-chief. The efforts of the first editor Ivan Milčetić, as well as the president of the Academy, Franjo Rački, to explain how external associates of the Academy should collect material, were only concretised by Radić in a text entitled Osnova za sabiranje i proučavanje građe o narodnom životu. The Osnova, sometimes referred to as the Questionnaire, is a methodological guideline for the collection of ethnographic material, conceived in three basic parts.<sup>32</sup>

<sup>&</sup>lt;sup>31</sup> The outcomes of the project will be available at the link: https://uciteljietnografi. omeka.net/.

<sup>&</sup>lt;sup>32</sup> Foundation for Collecting and Studying Materials on Folk Life consists of three main thematic units: questions in the first part discuss the impact of the environment on folk's life with chapters titled "Nature around Man", "Physical Structure of People",

Before Radić's ethnographic methodological text, various events shaped the later work on collecting ethnographic material within the Academy. On various occasions, Academy's president Franjo Rački talked about the importance of collecting knowledge about the life of the peasants. In those initial years, the Academy encouraged external collaborators to collect material and publish research with a focus on ethnographic content (cf. Čulinović Konstantinović 1979: 69). Students, public teachers, pastors, and chaplains, administrative officials, judges, lawyers, and nobles were invited to start collecting. The first structured invitation was published in the newspaper Friend of the People on 5 August 1897. At the same time, more work was being done on the structured and comprehensive instructions for collecting folklore and ethnographic materials. Non-structured ethnographic material kept arriving and was stored in the Academy's archive awaiting the review process and the editor's approval for publication. Finally, in the second issue of the journal Collection for Folk Life and Customs, Antun Radić takes over the editorship, publishes the text about ethnographic methodological guidelines and parallel with that works on the widening the network of collaborators who work intensively on collecting and summarising the ethnographic material. Vesna Čulinović Konstantinović notes that the main aim was to guide collectors on how to write an ethnography of rural areas (1979: 73). Reportedly there were 50 manuscripts stored in the Academy's archives up to 1898, and in 1900 about 36 manuscripts came from Croatia, Slavonia, Bosnia, Herzegovina, Dalmatia, Istria and Poland (74). Čulinović Konstatinović explained the importance of the Osnova for the early ethnologists' field work with following words "from the initial thematic study, Osnova encompasses the overall content of the people's life and culture, as they try to direct the recording to the monographic description of individual, mostly rural areas" (73).

In 1902, Radić left the position of editor of *Zbornik*, and Dragutin Boranić, who remained editor until 1954 together with Tomo Maretić

and "Language"; questions in the second part address how material culture shapes the everyday life of rural populations (chapters: "Life Necessities" and "Work"); in the third part, folk wisdom, mind, and heart become the focus and primary emphasis on which Radić concludes the story about the folk's life (chapters: "Life", "Law", "Customs", "Entertainment", "Poetry", "Beliefs", and "Experience, Knowledge and Wisdom").

came to replace him. Boranić held to Radić's idea that literate and educated participants who describe their own community should write down the ethnographic material, i.e. literate peasants or people from the countryside – students, priests, and teachers (Čapo 1997: 10). The emphasis that the material should be recorded by *amateur* collectors, although at the same time stakeholders of the same local community, implied certain advantages and disadvantages. The position of insiders enabled them to write thoroughly and continuously, while on the other hand, it was a position that did not allow them a certain distance from the developments in everyday life. Furthermore, the number of collaborators was minimal, which made it difficult to control the records, so the editors of the journal had the task of personally checking the records. Considering the number of records that arrived in a short time, this practice was soon abandoned or reduced to a minimum. From today's perspective, it is not possible to clarify how much of these records were prepared as idealised descriptions of everyday life, and if they were in tune with real-life situations and events. Despite this, the records remain an important source for contemporary researchers of everyday life in history as well as for the members of local communities. We can say that on many occasions they are the only witness to a time and way of life that passed.

In the journal Zbornik za narodni život i običaje, selected parts of the collected ethnographic material were subjected to linguistic analysis and were subsequently published. For example, the records of the teacher Luka Lukić from Brodska Varoš were published in several numbers of the journal (Number 24, 25, and 26) in 1919, 1924, and 1925, while his collection pertaining to the village of Klakar was published in two separate books, in 2016 and finally in 2020, under the title Luka Lukić: Description of the Village of Klakarja 1905 – 1953.<sup>33</sup>

<sup>&</sup>lt;sup>33</sup> Biography of Luka Lukić was published in 2016 under the title Luka Lukić: Teacher, Ethnographer and Melographer in Klakar, and in that year two books of his manuscripts were published. Until that time the manuscripts were kept in the HAZU Archive, and published under the title Description of the Village of Klakarja: Stories and Poems Collected from the Mid-19th Century to the 20th Century and Description of the Village of Klakarja: Construction, Furniture, Appliances and Tools at the Beginning of the 20th Century (Vanić 2022: 268). A new book entitled Luka Lukić: Description of the village of Klakarja 1905 – 1953 was published in the journal Zbornik za narodni život i običaje in 2020.

With this, we come to the contemporary context of work related to the Department of Ethnology of the Croatian Academy of Sciences and Arts (HAZU). Since 2002 the Department started publicising ethnological and folkloristic materials stored in their archive. In that way the manuscripts that have been stored in the Department's archive for more than a hundred years became available to the scientific and professional public as to the local community they were dedicated to. For example, in 2020, the manuscript of teacher Ivan Filakovac was transcribed and edited under the title Ivan Filakovac: Župa Retkovci 1898 - 1902 (cf. Dimšić 2022: 263). Despite all the efforts and work that the Academy's employees and associates invested in the preparation of the manuscripts, one part of them remains stored in the archives. Activities of the UČeka project tried to focus on these manuscripts in order to open them to the public. In the scope of the project, we developed activities such as a digital interactive map. The interactive map, among other things, indicates the frequency and regional prevalence of a teacher's ethnographic records. The map, as well as the project's digital platform on omeka.net, was generated to attract new users for the specific content. The digital interactive map for the moment contains the record of the teacher's manuscripts from the Archives of the Department of Ethnology of the Croatian Academy of Sciences and Arts. Namely, the manuscripts of the following teachers: Stipe (Stjepan) Banović from Zaostrog, Pavlina Bogdan Bijelić from Cavtat, Nike Balarin from Konavle region, Bogdan Krčmarić from Smiljan, Juraj Milaković from Petrinja, Pavao Wolfi from Vinica, Rade Bosnić from Banovina region, Stjepan Žiža from Istrian peninsula and Štefanija Bernas-Bilošević from Žumberak region. The manuscripts were reviewed giving special attention to the chapter named "School". In Osnova, questions about school are included in the third part as the last chapter, and it contains nine questions (Radić 1929: 45). In our first reviews of the manuscript, it is evident that the teachers were not dedicated to just answer questions about everyday school life. Throughout all the manuscripts we can read about the attitude of local people towards the school and education in general. Parts of the manuscripts describe the status of children in the family, the children's labour, in general, and the context of the time that tells us more about the parent's attitude towards school i.e., the importance that education had in communities that depended

on physical work of all its members. The records analysis included insight into teachers' views on the living culture and everyday life, also an insight into the positioning of teachers in relation to the communities in which they worked.

# THE MANUSCRIPT OF STJEPAN AUGUST: BETWEEN PUBLICATION AND DIGITISATION

Our research in the State Archives in Bjelovar revealed an interesting manuscript Village and Villagers written by teacher Stjepan August. For more than a hundred years, the manuscript was stored first in the local school archive, and, since 2018, it has been stored in the State Archives in Bjelovar. Considering the ethnographic material it contained, it was decided to prepare it for publishing. The manuscript was written in a simple notebook with the intention to describe the village and its inhabitants, rather differently but also with some elements similar to the school memory books known at the time.<sup>34</sup> The manuscript is dated with two dates, at the end of the "Preface" it is 5 July 1905, but in other parts of the manuscript, the dates are between 1902 and 1914. It can be assumed that the author began to write down his notes in the notebook at some point (possibly in 1905, when he was writing the "Preface") and listed the dates as he wrote other parts of the text.

Considering the different quality of the handwriting some parts of the manuscript were written directly into the notebook, while others were copied from another one. Stjepan August's manuscript is a combination of a school memory book, containing information about school life which the author supplemented more extensively with information about the everyday life of the village. That fact prompts us to think about whether the author, although he does not mention it any-

<sup>&</sup>lt;sup>34</sup> The order on how to write memory books was passed in 1880. The memory books were supposed to have two parts: the first part refers to the history of the specific school, while the second part is imagined as a yearbook of the school. Aladrović Mehandžija raises the question of the reliability of data that was recorded in the historical part of the memory book because the teachers were inclined or didn't have access to official documentation, so they were forced to rely on "oral transmission" of knowledge among the teachers and data that the villagers passed on to them (2012: 50).

where, was familiar with Antun Radić's methodological text. Probably inspired by the school memory books, he decided on a more complex data recording. In the preface of the manuscript, Stjepan August addresses his successor in the position of the head teacher as "dear comrade and successor", and announces his intention with which he wrote the manuscript, "so that after reading it, you will know where you live, and accordingly, you will shape your life for the benefit of the Croatian school and our dear homeland" (August 1905). Stjepan August writes from his own experience and advises his future successor on how to gain the trust of the villagers and how to get to know "their essence". For this, he must have interacted intensively, not only with the children who attended school but with all members of the community, in situations where he met their strengths and weaknesses. This is how the villagers will accept the teacher who is a stranger, "as the father of their children, they will recognize you as their true friend, and they will be happy to listen to your advice, fulfill your wishes, and you will be a satisfied Croatian public teacher because you will achieve great success in your work for religion and homeland" (August 1905). The manuscript Village and Villagers is divided into three parts. In the first, titled "Description", the author records the daily life of the village, according to his own view, but also through the stories of the villagers; in the second part, titled "Biography" (Croatian Živopis), he completely independently describes his view of the inhabitants of the village, portraying in detail individual families, their negative attitude towards children's schooling in the Croatian language, and in the same context he also mentions the hygiene of the household and the attitude towards learning the Croatian language; in the third part, "School Work and Success/Addendum", dedicated to his work at the Blagorodovac school, he records his own daily work setting and includes facts about the school's history.

First, we started with the transcription of the manuscript. When this was finished, we proceeded to a more detailed review of the manuscript and confirmation of the relevance of the material it contains. Teacher August's record is filled with stories about local everyday life, historical figures and references to various historical events that shaped the everyday life of the population, from immigration, and interaction of settlers and domiciled population to the development of crafts and agriculture. The role of the nobles and local dignitaries connected to the Austrian administration, the arrival of the colonists, and their taking over administrative tasks greatly dynamised the local everyday life and the affairs of the local population. The author's opinion, remarks, criticisms and praises of the local way of life can be found in the manuscript. The strong voice of the author who does not refrain from presenting all the problems caused by immigration, cultural praxis, daily habits, etc. that characterise the local population permeates the entire manuscript.

The process of rewriting lasted a little over six months, considering the difficulty of reading the author's handwriting, as well as managing the terms that are no longer in use. Also, due to the author's tendency to use Germanisms, written down in literal translation, it was rather difficult to get through the sentences. The transcription was checked twice more, first with the simultaneous reading of the original record and once independently. In the second phase, critical processing of the manuscript followed as preparation for publication preceded. This part entailed writing an introductory study, which explains all the steps of preparing the manuscript for publication but also contextualises the manuscript itself so that today's readers can better navigate the text. So the methodological design of the manuscript was divided into three steps, firstly focusing on content editing. We took an interdisciplinary approach by reading different literature and sources in order to show as clearly as possible the context in which the manuscript was created. Part of the literature was related to enlightening the teacher's role and the state of education more than a hundred years ago. In this first step, we were guided by preparing the entire content of the manuscript so that it would be comprehensible to a modern reader. The second step required the editors to employ a sharp and critical eye as we meticulously examined whether the author expresses himself clearly in the text, identified any contradictions, and determined how to approach them. We standardised punctuation, the use of abbreviations, and corrected spelling mistakes, while also working on improving the author's illustrations. Following this, we checked the alignment of chapters, sub-headings, and citations (including the use of single or double quotation marks), and standardised scientific nomenclature and terminology. The third step was the most intensive, as it involved checking for consistency both on the part of the author and on our part as editors of the manuscript. Any inconsistency could raise doubts about the accuracy and thoroughness of the facts and that alone can make the relevance of the prepared manuscript questionable. While the *Village and Villagers* manuscript is being prepared for printing, the scanned parts of the manuscript, together with the documents, have been uploaded to a digital platform, marking the beginning of "their life in the virtual world".

# THE PLATFORM OMEKA.NET TAILORED FOR THE UČEKA PROJECT

Participants in the UČeka project researched various aspects of the topic using archival material. Each participant focused on an individual institution and funds, collecting information regardless of its direct relevance to their specific topic. This approach was logical as we intended to share all collected materials, thus saving each participant from the need to visit multiple archives to access the same documents. To meet the project's requirements, we acquired equipment capable of high-quality recording of archival material. All participants were familiarised with the *Instructions for the Digitization of Documentary and Archival Material* published by the Croatian State Archives at the beginning of 2021. This facilitated the sharing of high-quality documents among project participants. By seeking approval for the use of the material, we ensured that the collected materials can be presented in a digital environment.

During the planning of the project tasks, a virtual exhibition was foreseen, which would point out the various values and possible applications of the collected material in a concise and visually interesting format. Therefore, it was necessary to come up with a way of displaying and sharing the collected archival material.

Considering the diversity of the material and previous experience of the State Archives in Bjelovar, with the fact that the project does not provide funds for creating a database, Omeka.net was chosen. The Omeka digital platform was launched by the Roy Rosenzweig Center for History and New Media and since 2009 has been managed by the

non-profit Corporation for Digital Scholarship (Omeka.org 2022). It is an open-source content management system that allows users to use software plugins without much coding knowledge, thus achieving diversity and a greater level of customisation of the platform to users (Ceja Alcalá and Alaniz 2018: 164). It is used by many institutions of the GLAM community mostly in the USA. In addition to Omeka.net, it is possible to choose Omeka S and Omeka Classic, but they require storage on the user's server and certain system requirements that were not available to us when planning the project. Omeka.net offers hosting for the publication of digital collections and exhibits, relieving users of the need to worry about installation or hosting. It is free for a plan that includes 500 MB of storage space. It uses the Dublin Core schema for metadata and supports multiple file formats. The interface is customizable and has several different themes that can be used in the trial version. It also has pre-installed plugins that can be customised. It enables the hierarchical arrangement of document collections and the creation of an exhibition, which was important to us for the presentation of materials on this project (Omeka.net 2022). After creating an account on Omeka.net, the administrator can choose one of two themes and configure them, install plugins and set language preferences (Croatian is available). On the Dashboard, there is an administrative overview of basic information about recently added items, and collections, and the total number of all added items, collections, tags, plugins, exhibits and chosen themes of the pages. The administrators' page contains tabs that navigate through Items, Collections, Item Type, Tags, Exhibits and Simple Pages section and activated plugins. The Collections tab is used for grouping added items by theme. After assigning items to the collection, you can create a visual layout along with the text commenting on the items of the exhibition. Simple pages are used for adding a homepage for your site and writing about the project in pages subordinate to the homepage. These pages can be organised hierarchically, which enables the grouping of information according to the needs of the project. In due time, all material collected within the project will be posted on Omeka.net.

By organising the pages, it is possible to group the material into collections that are made on the principle of respecting the provenance of the source. The next criterion is the grouping of the collected materials into sections, and in the case of a larger amount of material, it will be divided into smaller collections and described by a basic set of metadata. In this way, easier navigation through a large amount of collected materials will be achieved. The ability to choose the name of the page according to our wishes, another user-friendly option of Omeka.net, helped us to make the project more visible. Although the acronym of the project was an obvious choice, we called the page https://uciteljietnografi.omeka.net/ because it highlights the important keywords of the project - teachers and ethnographers. Currently, there are six collections on the project platform - Ethnographic memory book documents and the teacher Stjepan August's manuscript about the village of Blagorodovac divided into five collections. Although it is really a single-bound manuscript, the digitised material is placed on the platform in five logical units, i.e. the way the author divided his manuscript. Each page is loaded individually and assigned a place in each collection. Blank manuscript pages were not included because we were focused on information that might be of interest to users, and blank pages would make it difficult to read and browse through the collection. The collection named Ethnographic memory books contains documents created in the period from 1939 to 1945. Ministries and offices that were in charge of contacting the schools were sending instructions regarding the creation of Ethnographic memory books. Each individual scan can be assigned a set of metadata according to the Dublin Core scheme and tags, i.e., keywords, which makes it easier to search and find what the site visitor is interested in. The participants' materials, collected from other archives, will be placed on Omeka in the next phase of our work. They will be grouped into thematic collections and equipped with keywords and metadata. After that, a virtual exhibition will be set up in which the collected material will be interpreted and presented to the audience.

### CONCLUDING REMARKS

The reason for using a digital platform is the need to ensure the wider availability of collected materials. It is important to emphasise that in this way, users do not have to download additional materials or applications to their computers and mobile phones, thus putting an additional load on their devices. The entire content is available on the platform and can be discovered at one's own pace. Also, in situations where the user is not familiar with the subject matter or the work of individual ethnographer teachers, a simple search of websites by keywords will bring up documents that contain the requested information.

We tried working with the advantages of omeka.net in order to reduce the users' workload in searching for what is important and interesting to them. Also, the layering of the platform makes it easier for the user to navigate in search of information. Every user chooses what he is interested in at a given moment: collections, exhibitions, documents or searching the content with the help of keywords.

The project's intention is to continuously supplement the platform with relevant material. That will ultimately form a unique base of collected material of teachers as creators of ethnographic knowledge. The link to Omeka will be posted on the websites of the project participants' institutions: the University of Zadar and the State Archives in Bjelovar, and will be offered to the institutions whose material it contains. The material collected in this way will be available to the academic community, both professors and students, researchers, as well as anyone interested in the topic and material.

#### **REFERENCES LIST**

- Aladrović Mehandžija, Katarina. 2015. "Najstarije školske spomenice Broda i okolice iz Državnog arhiva u Slavonskom brodu". Anali za povijest odgoja 14: 49–68.
- Ceja Alcalá, Janet and Desiree Alaniz. 2018. "Toward Community Engaged Archiving: Building a Digi-Rasquache Archives". In *Cultural Heritage Care and Management: Theory and Practice*. Cecilia Lizama Salvatore, ed. Maryland: Rowman & Littlefield Publishers, 153–169.
- Čapo Žmegač, Jasna. 1997. "Antun Radić i suvremena etnološka istraživanja". Narodna umjetnost 34 (2): 9–33.
- Čulinović Konstantinović, Vesna. 1979. "Historijski prikaz etnoloških istraživanja i interesa za narodnu kulturu u Hrvatskoj". Etnološka tribina 2: 67–88.
- Dimšić, Katarina. 2022. "Ivan Filakovac, Župa Retkovci 1898.–1902., Zagreb-Vinkovci 2020". Rad Hrvatske akademije znanosti i umjetnosti. Razred za društvene znanosti 550=56: 263–265.

- Markovac, Marijan. 1940. Etnografske spomenice i etnografske zbirke u narodnim školama Banovine Hrvatske. Zagreb: self-published.
- Prpić, Lara and Marija Fajdić. 2021. "Školske spomenice kao arhivsko vrelo za proučavanje regionalne povijesti i povijesti školstva Like i okolice". *MemorabiLika* 4 (1): 119–146.
- Radić, Antun. 1929. Osnova za sabiranje i proučavanje građe o narodnom životu. (Drugo izdanje). Zagreb: JAZU.
- Vranić, Dunja. 2022. "Luka Lukić, Opis sela Klakarja 1905. 1953., Zagreb Klakar, 2020". Rad Hrvatske akademije znanosti i umjetnosti. Razred za društvene znanosti 550=56: 267–274.
- Župan, Dinko. 2010. "'Živio svijetli ban!' Kako je u školskim spomenicama i izvještajima zabilježen posjet bana Dragutina Kuhena-Hederavaryja nekim mjestima Virovitičke županije 1889. i 1893". Scrinia Slavonica 10: 212–235.

### LIST OF SOURCES

- August, Stjepan. 1905. "Blagorodovac Elementary School". In Village and Villagers. Bjelovar: State Archives in Bjelovar (manuscript, HR-DABJ 842).
- Omeka.net. Available at: https://www.omeka.net/ (accessed 5 December 2022).
- Učitelji kao stvaraoci etnografskog znanja. Available at: https://uciteljietnografi.omeka.net/ (accessed 5 December 2022).

## ADAPTING ENGLISH-LANGUAGE PEDAGOGY TO CULTURAL HERITAGE AND DIGITAL HUMANITIES IN FRANCE: RESULTS FROM USING THE IIIF ONLINE WORKSHOP WITH FUTURE RESEARCH ENGINEERS, LIBRARIANS, AND ARCHIVISTS

### EDWARD J. GRAY

DARIAH-EU / Huma-Num (CNRS) Paris, France edward.gray@dariah.eu https://orcid.org/0000-0002-5201-1014

The preservation of cultural heritage is one of the fundamental missions of L'École nationale des chartes in Paris. For over fifteen years, the École has also offered a Masters entitled Technologies numériques appliquées à l'histoire (TNAH), or Digital Technologies Applied to History. This program trains future archivists, librarians, and research engineers in a common program that focuses on the technical aspects of the preservation and valorization of cultural heritage. Despite this common program, or perhaps because of it, graduates find careers all across this spectrum. It is thus a challenge to find subject matters that are transversal to all these potential interests. For three years, as a chargé de cours (lecturer) for the English-language portion of the curriculum, I have chosen to use the IIIF Online Workshop (https://training.iiif.io/iiif-online-workshop/index.html). This training workshop takes students through the process and logic of the IIIF standard, from understanding its uses to manipulating the image API, and finally to creating and hosting one's annotations. It finishes with a project, chosen by students, where they upload, manipulate, and annotate their images, which they then present to their peers as a "proof of concept" for a theoretical IIIF project at their imagined institution. As a whole, this IIIF assignment permits not only an assessment of reading and spoken comprehension but also oral and written communication, alongside a technical standard that is increasingly more in demand.

Keywords: pedagogy, IIIF, English as a second language, digital humanities, cultural heritage

With respect to the importance of multilingualism, mastery of English is a necessary skill for digital humanists.<sup>35</sup> It is used in documentation, training, and professional exchanges between colleagues. Just as standards lead to interoperability for data, so does a single language allow for researchers from different backgrounds to communicate and exchange their ideas. Thus, it is commonplace for Digital Humanities degrees to include some instruction on or in English, and the *Technologies numériques appliquées à l'histoire* Master's program at L'École nationale des chartes in Paris is no different. As the instructor of this class, I wish to present the results of three different promotions of students (2021, 2022, and 2023), so that others can understand how one may surmount the combined challenge of providing effective English-language instruction on digital humanities content, including relevant and practical use cases for the use of English, all within a few hours of class time.

The solution adopted to confront this challenge and provide the most useful course was to adopt units and assignments geared towards professional use cases, the keystone of which is the adaptation of an existing training seminar on IIIF. Taught in English and aimed at GLAM professionals, the IIIF Online Workshop provides an ideal vector for introducing students to a DH-relevant professional experience, as it allows the course to focus not on English as a language, but on the use

<sup>&</sup>lt;sup>35</sup> I wish to extend my deepest thanks to Glen Robson for his tireless work in advancing IIIF and especially his work in developing the IIIF Online Workshop, as well as some of the tools he specifically developed to help make this course, and IIIF, more accessible to students both of his own training session and others. I also wish to thank the three cohorts of the TNAH Master's degree, 2021, 2022, and 2023, who took this course, which challenged them not only linguistically but also technically. Finally, I wish to thank Koraljka Kuzman Šlogar and the entire DARIAH-HR team for their work in running the Digital Humanities & Heritage international conference and for the opportunity to share my work with others.

of English to complete a technical task. Throughout the four units in the workshop, students learn what is IIIF and why it is used, how to master the Image API, the Presentation API, and finally how to make and publish annotations. Students are asked to complete the workshop as if they were enrolled in it as part of their professional duties. This roleplaying aspect asks students to reflect not only on *how* to understand and deploy IIIF in their fictive institution but also on how to explain (via a presentation) the added value of implementing IIIF to their colleagues.

In terms of pedagogical outcomes, this approach leads to a holistic familiarization with the English language and familiarization with the IIIF standard and APIs more generally for the students. They are confronted with written and oral comprehension activities, in the process of following along with the workshop units. Additionally, students must compose in English to create their IIIF manifests and the accompanying annotations, just as they are required to speak English in their mock presentation at the end of the workshop. As a whole, then, this capstone project asks students to use both comprehension and composition skills. Beyond the English-language aspects, students are also expected to gain a working familiarity with APIs in general and the IIIF standard in particular, understanding both why it exists and how it functions. Students define a coherent scientific project around which they feature images with accompanying annotations. They are tasked with using the IIIF Online Workshop to host their own, IIIF-compliant images, create a canvas bringing these images together, and provide annotations that contribute to their scientific project. This capstone thus engages in one of the major missions of the TNAH Master's program, which trains students to be effective bridges between the traditional worlds of GLAM and research institutions and the technical skills and opportunities offered by digital humanities.

The TNAH Master's program began at L'École nationale des chartes in 2006, and since then has been one of the premier digital humanities Master's programs in France. That L'École nationale des chartes, which has its origins in 1821, is at the forefront of digital technology should not be surprising. The institution is a grande école specializing in the preservation and study of cultural heritage and has a long and storied history in advancing archival and library sciences, as well as the study of history. The TNAH is thus a continuation of the traditional focus on the preservation of cultural heritage, adapted for a new, digital age. Alumni go on to become archivists, librarians, researchers, and engineers in both public service and the private sphere; and it is important that the pedagogical choices taken by instructors are applicable to these fields.

The TNAH Master's degree lasts two years, with the first year focused on the study of sources, with a short introduction to technical matters such as the basics of programming or XML. This serves as a basis upon which the fast-paced and technically driven second year is built. The first "semester" of the second year is dedicated to intensive technical training, lasting seven months, from the end of September until mid-March. After a period for working on final projects, students begin a three- to four-month internship at a GLAM, research, or private institution during which they will put their digital humanities skills to the test in an institutional context, with a technical deliverable to hand in by the end of the internship. After these internships, students write a Master's Thesis that problematizes the work that they completed, which is then defended in front of a jury before completing the program.

What then, is the place of English in this curriculum? If students are free to take any of the "living languages" offered by the École in their first year, in the second year they are all required to take the Anglais: Langue de l'informatique (English: Language of Computer Science) class. As the instructor, I felt obliged to ensure that this class was useful and productive for our students, no matter what career path they chose. As a certain level of English mastery is a prerequisite for the second year of the Master's program, I decided that the class would not contain simple reviews of grammar or vocabulary, but would instead focus on practical, daily uses of English that the students could encounter in their professional lives. Lessons vary from best practices on how to present oneself, such as a short bio for a conference, to general English-language academic etiquette, to lectures and discussions on research infrastructures, the European research projects ecosystem, or the ethical challenges of using GAFAM technologies. I chose the IIIF capstone project as a way to not only ensure a wide-ranged mastery of English but also because I had noticed a key need for IIIF familiarity in the job postings relevant to our future alumni.

The International Image Interoperability Framework, or IIIF, is a powerful, open-source standard that allows patrimonial and research institutions to easily share high-quality image-based resources without being burdensome on technical capacity. It is a solution to the thorny problem of how an institution can provide high-quality media to its stakeholders, without the heavy financial and environmental costs related to transmitting such massive amounts of data. Institutions host their images on an IIIF-compliant server, which allows them to be retrieved using the IIIF Image API to an IIIF viewer, hosted elsewhere. In essence, when the user zooms onto an image, the API will only request the requested number of pixels of that image and in the requisite quality. This provides for a smoother viewing experience, that is friendly to both the end-user and the institution. Beyond this Image API, IIIF also has a Presentation API, which allows for the institution to provide metadata, structural organization of a collection of images, and even annotations via an IIIF Manifest, which is itself a JSON file. This manifest includes a host of metadata, such as labels, licenses, attribution, and a sequence of images. More importantly, it defines canvases, upon which are "painted" the images and, if necessary, any accompanying annotations on those images. While the Image API and the Presentation API are the two main IIIF APIs covered in the IIIF Online Workshop (and my course), there are other APIs that allow for content search, authentication, and georeferencing. For all these reasons, IIIF is a key technology for the students of the TNAH Master's degree to understand, and is useful for their future careers, whether they find themselves in research or cultural heritage institutions.

A final key component to understanding IIIF, and one that I stress to my students, is the IIIF community. The IIIF standard is at its very core about open science and breaking down silos to allow for easier sharing of image-based resources. The community of practitioners and developers that maintain, implement, and continue development on these standards is just as open and welcoming as the code they write. It is a truly collaborative effort, and one that represents the best of digital humanities; which is another reason why I made the choice to focus on this technology. The proof is that the IIIF Online Workshop developed by Glen Robson and the IIIF team in 2020, while provided as a paid training session taught and supported by IIIF staff, is also available online for anyone to use at any time and at their own pace.

The IIIF Online Workshop is hosted via GitHub pages and includes five sections, each corresponding with one of the days of the training course. The first section, IIIF Basics, gives students an introduction to IIIF, why it exists, who uses it, how it works, and the overall value proposition of adopting the standard. Like with all the sections of the course, the IIIF Basics combines text with images and embedded video presentations. Additionally, GitHub pages allow for embedded widgets, which give practical experience with IIIF tools, without having to open a new page. Thus, in the introduction, students can manipulate IIIF images via an embedded Mirador viewer, in this case using Biblissima's reconstruction of a damaged manuscript from Châteauroux, in central France. These illuminations, from Grandes Chroniques de France, were cut out of the manuscript at some point in the nineteenth century, and are today found at the Bibliothèque nationale de France (National Library of France). With this exercise, students see how IIIF can reconstruct and repair manuscripts that otherwise are held in separate institutions, and it is a poignant example of the capacity of IIIF to help advance patrimonial collections.

The second section, Image API, leads off with an hour-long video (a recording of Glen Robson's presentation to a previous training session) presenting the IIIF Image API. It is followed by a step-by-step text that takes students through the basics of the IIIF Image API, with another plugin that allows students to manipulate the URL of an IIIF image to show, concretely, how changing the URL can interact with the API.

The lesson on the Image API ends with asking students to select and upload images into an IIIF-compliant web server. Students are given several options, including hosting on Internet Archive, using the IIIF Workbench, or setting up and running one's own Cantaloupe IIIF Server. While the latter is better geared to those who work for an institution and are interested in pursuing a more structurally-sound and long-term option for hosting IIIF images, the former two options are best suited for the Master's students in DH as they are simple and easy to deploy. I require my students to use the IIIF Workbench, which hosts images via GitHub pages and is a specific, tailor-made cloud service developed by Glen Robson for the IIIF Online Workshop. Students thus have continuous access to and control over their images, as they are on their own personal GitHub, and can continue to valorize their work well after the end of the class, without relying on the third-party solution offered by the Internet Archive or having to go through the hassle of setting up and maintain an IIIF server with Cantaloupe. This option comes at the cost of having level-0 compliance with IIIF, that is, that it will work with viewers but the tiles upon which one can zoom are predefined and not as responsive as with a true IIIF server, nevertheless, this solution is more than adequate for DH Master's students.

The next section focuses on the Presentation API, which is where students begin the real technical challenge of the IIIF Online Workshop. Throughout this lesson, again introduced by an hour-long video presentation of Glen Robson and accompanied by a step-by-step guide on how to complete tasks, students will take the IIIF images they have already uploaded in the previous lesson, create an IIIF manifest from these images, and complete that manifest with appropriate metadata. Students may select any images they like, either their own (which can be uploaded into an IIIF server using the methods described at the end of the previous lesson) or those from existing IIIF-compliant repositories; it is up to them to define the unifying theme behind their manifest, which forms the core of their capstone project. Students are guided on how to use the Bodleian Manifest Editor to create an IIIF manifest. The next step is to create an environment that will allow students to edit their JSON manifest, as well as view the manifest via a web server.

The IIIF Online Workshop guides students through the process of manually editing the manifest, through which students will gain deeper familiarity with the mechanics of a process that is usually automatic. This is particularly valuable in a pedagogical setting, as it is by manipulating these mechanics that students gain an appreciation for the inner workings of IIIF. As VS Code, the open-source editor recommended by the IIIF Online Training allows for local hosting, students can save their manifest and check to see how the changes they have added are reflected in IIIF viewers, thus seeing step-by-step how their manifests evolve and the direct impact of each change they make. Students are then directed to publish their manifest through the IIIF Workbench, which allows students to have a publicly available link (as opposed to the locally hosted link they used earlier in this lesson) to their work, which can then be used in a variety of IIIF viewers, such as Universal Viewer or Mirador. At the end of this third section, students should have a publicly-available IIIF manifest, with hosted images, ready for annotation.

The fourth section of the IIIF Online Workshop is a direct continuation of the previous lesson and expands upon the functionalities of the Presentation API to guide students in making their own annotations. These annotations are to be written in English, which engages students with the English-language portion of this course. Students first complete some exercises in annotation with Mirador.<sup>36</sup> These annotations, which are only stored in the browser cache and therefore are not permanent, give students a low-stakes opportunity to learn the basics of annotating before moving on to more permanent solutions. Once again, the choice here is to privilege a step-by-step process that allows students to use IIIF in a very detailed way to ensure that they understand the effects of future, more automated processes that they will encounter in their professional lives. After a discussion of multiple annotation options available to students, the IIIF Online Workshop guides students step-by-step with the SimpleAnnotationServer, developed by Glen Robson during his time at the National Library of Wales. Students are directed to write and save annotations for their own manifests by loading their manifest into the Simple Annotation Server, creating their annotations, and then finally downloading their annotations from the SAS to then re-upload into the IIIF Workbench. This process creates a publicly-available annotation of their IIIF manifest, which allows not only for me to grade their work, but also for them to be able to valorize their project and show prospective employers that they are familiar with IIIF. In the course of this lesson, students are also directed on how to add search functionality to their annotations. At the end of the lesson, Robson gives resources to more robust, but technically advanced,

 $<sup>^{\</sup>rm 36}$  The current edition of IIIF Online Workshop references Mirador 2, but will soon be updated to Mirador 3.

annotation server options, which students can reference for future projects.

The fifth and final section of the IIIF Online Workshop, both for the official IIIF Online Training seminar and my own adaptation, is a presentation of the work done by the students. Throughout this capstone project, I ask my students to roleplay, pretending that they are employed by a cultural heritage institution or research lab and that they are sent to take the IIIF Online Workshop in preparation for a potential switch to IIIF-based infrastructure in their institution. Therefore, students are to take notes, use part of their "institution's" collections as a sample, and report feedback to their colleagues and superiors about IIIF and what it has to offer to their institution. The ten-minute presentation at the end of the course, then, is meant to fulfill the feedback and reporting part of the exercise, asking students to not only summarize what they did but also to consider the technical requirements and the strategic benefits that IIIF can offer a GLAM or research institution. Students have risen to this challenge, and in the nearly sixty presentations I have seen over the last three years, I have seen remarkable engagement with the basic roleplaying premise. Students have reported that allowing them to select their own images allows them to become more invested, and in some cases, continue previous or ongoing research projects.37

My guiding principle in adapting this resource to my class was to present students with as close to a real-life experience as possible. My inspiration came from my own experience with the paid version of this course in September 2020, taught by Glen Robson. Indeed, while my students will most likely spend most of their professional lives working in French, it is very likely that they will at some point take training in English, and almost guaranteed that they will need to read technical specs or documentation as well as interact with tools that use English in the interface. This course, thus, represents a true-to-life experience. There are several differences, however: notably, rather than everything being done in the space of a week, the nature of the academic calendar

<sup>&</sup>lt;sup>37</sup> It is not uncommon for students to have completed masters-level or even doctoral work before taking the TNAH Master's, and each year there are several students that continue to engage with their research interests in the IIIF capstone project.

for the second year of the TNAH means that classes are once a week and thus the IIIF capstone project is spread out over several weeks. Early in the semester, I introduce the assignment during class, which includes an introduction to IIIF as well as an explanation of the assignment, with an opportunity for questions. Students are then mostly on their own, to follow the IIIF Online Workshop at their own pace, though students are asked during each of the following classes how things are going. This relative independence is a change from Glen Robson's method, but it is a conscientious choice on my behalf to force the students to interact with the English-language resources, both textual and audiovisual, that make up the IIIF Online Workshop. It represents a real-to-life experience of how one must use available documentation, or one's network, to help find the answer to questions. Despite this choice, I have found it useful to have periodic check-ins and options for students to ask questions during the semester to ensure that they are not feeling left behind.

This approach is not without potential flaws, but I think that they are possible to overcome and lead to reflections that are valuable for these young professionals. In light of procrastination, which is eminently understandable given the heavily loaded semester, I have found good success in having students submit a project proposal early on in the semester, to benchmark their progress and ensure that they understand what they are doing and not wait until the last minute to begin the assignment. Indeed, the greatest frustration that students have found is that this project truly does take some time to complete (I estimate between eight to ten hours, based on one's familiarity with English). The assignment simply cannot be done in the space of an afternoon or a weekend. This frustration can be further compounded by the reliance upon multiple tools - and thus multiple points for failure - that are maintained by IIIF (Workbench, SAS) for their own version of this workshop. Sometimes, for entirely understandable reasons, these tools are temporarily down, or there is a setting in the student's own set-up that leads to incompatibilities or failures. For instance, students should ensure that they have a rapid, stable internet connection when uploading images and annotations in order to avoid problems. On more than one occasion a student has run into a crisis because they waited too long to begin the assignment and had to rush through only to find that an important tool was down for maintenance. It should be understood that I do not say this to critique the excellent work of Robson and the IIIF team who develop and maintain these tools, but only to let others who wish to use this method understand the potential risks involved. Indeed, I think it is a *valuable* experience for our graduates, as it can lead to reflections on the benefits and costs of relying on software or tools that are not hosted in one's own institution or are less likely to be fixed immediately. This represents the reality of using open-source software and decentralized services and it is an important lesson as digital humanities makes great use of these types of tools and services.

In conclusion, this workshop, originally developed by IIIF Technical Coordinator Glen Robson and adapted to the needs of the TNAH program, has multiple benefits for the students. Above all, it demonstrates a real use-case in using the English language - attending a training seminar - in the course of one's professional life, which corresponds to the vocation of the TNAH Masters to prepare its students for their future careers. It also applies broadly to all future career choices of the students, being just as relevant to an archivist as a research engineer in a humanities lab. As well, it gives real, practical experience with APIs - so that students understand intimately how this fundamental technology works. Finally, allowing students to choose their own subject for the final project enables them to embrace their interests and creativity, thus increasing their engagement with the assignment. Indeed, I feel that asking students to choose their own subject matter for the final assignment increases their investment and intrinsic interest in the assignment, which has a benefit not only in the strictly pedagogical sense as they are more engaged in the final project but also in the longer-term perspective of providing them with a deliverable that they can use to demonstrate their skills. From a language-learning perspective, the course is pedagogically sound, as it confronts students not just with reading and spoken comprehension, but also with written and oral composition with the final project and the presentation to the class. The IIIF Online Workshop is thus a powerful tool for Englishlanguage pedagogy in digital humanities and the preservation of cultural heritage.

### **REFERENCES LIST**

- Bodleian IIIF Manifest Editor. Available at: https://github.com/bodleian/ iiif-manifest-editor (accessed 2 May 2023).
- Microsoft. 2015. Visual Studio Code. Available at: https://code.visualstudio. com/ (accessed 2 May 2023).
- Mirador Viewer. Version 3. Available at: https://github.com/projectmirador/mirador (accessed 2 May 2023).
- Robson, Glen. 2020. IIIF Online Workshop. Available at: https://training.iiif. io/iiif-online-workshop/ (accessed 2 May 2023).
- Robson, Glen. 2022. IIIF Workbench. Available at: https://workbench.gdmrdigital.com/index.xhtml (accessed 2 May 2023).
- Robson, Glen. Simple Annotation Server. Available at: https://github.com/glenrobson/SimpleAnnotationServer (accessed 2 May 2023).
- Universal Viewer. Available at: https://github.com/universalviewer/universalviewer (accessed 2 May 2023).
# **COMMUNITY VERSUS CORPUS**: ACCESSIBLE LANGUAGE RESOURCES FOR INDIGENOUS SPEAKER COMMUNITIES AS A WAY TO PRESERVE AND PROMOTE CULTURAL HERITAGE

# ELENA LAZARENKO

University of Hamburg, Institute of Finno-Ugric/Uralic Studies Hamburg, Germany elena.lazarenko@uni-hamburg.de

# ALEKSANDR RIAPOSOV

University of Hamburg, Institute of Finno-Ugric/Uralic Studies Hamburg, Germany aleksandr.riaposov@uni-hamburg.de

This paper explores the ways the long-term project INEL strives to adapt linguistic corpora to be used as a web resource by the native speakers of endangered Northern Eurasian languages. Background motivation, concept, and choice of materials and tools are discussed. A work in progress on three language pages – Dolgan, Nganasan, and Selkup – as well as on additional interactive materials is presented.

Keywords: speaker community, indigenous languages, corpus linguistics, online resources, language preservation

## INTRODUCTION

From over 7,000 living languages of the world, only a few dozen are well represented online. While even some of the languages considered "safe" according to the UNESCO criteria often lack web resources, the situation with endangered languages is quite alarming. It might be hard to find well-formed and easy-to-use web resources for indigenous speakers. Even though the people working in the fields of language documentation and corpus linguistics do their best to gather language data and preserve it in a sustainable manner, many of these initiatives end up being out of reach for non-linguists and – most importantly – for the speaker communities. This happens for a multitude of reasons, but the resulting incongruence between what language resources have on offer and what speaker communities would have liked to use ultimately does not help much to achieve language preservation goals.

In this paper, we are going to present the approach the project INEL undertakes in order to bridge this gap by creating easily accessible online materials for the speaker communities of endangered Siberian languages, based on the language corpora developed in the project. In the following sections, we will outline what the project is about, discuss the common issues regarding endangered language visibility on the web, and present our concept of turning highly specific corpora into resources for community speakers.

#### **1. THE INEL PROJECT**

The long-term project INEL ("Grammar, Corpora, Language Technologies for Indigenous Northern Eurasian Languages")<sup>38</sup> started back in 2016 and has ever since been working with the linguistic data from highly endangered languages and language varieties that stem from the Northern Eurasian Area. Following predefined language-specific workflows, INEL acquires, annotates, digitally curates, and publishes language corpora and associated resources. Up to date, corpora in such languages as Kamas (< Samoyedic < Uralic), Dolgan (< Turkic), Evenki (< Tungusic), and Selkup (< Samoyedic < Uralic) have been made publicly available under open access conditions (Lehmberg 2022: 71). As of 2022, there are two more language corpora in development, namely Nganasan (< Samoyedic < Uralic) and Nenets (< Samoyedic < Uralic).

<sup>&</sup>lt;sup>38</sup> The project is funded by the German Federal Government and the Federal States in the Academies' Program, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Program is coordinated by the Union of the German Academies of Sciences and Humanities.

The standards of data preparation of INEL corpora are high and are based on sustainability and long-term availability, and this limits the data formats in use. The final versions of INEL corpora, having travelled through various tools and formats during the pre-processing stages, are delivered in XML-based markup format of the crossplatform application framework EXMARaLDA<sup>39</sup> (Schmidt and Wörner 2014). EXMARaLDA tools help to present corpus materials sustainably and consistently, creating an easily navigable environment of corpus metadata, transcriptions, linguistic annotations, and, if applicable, audio and video sources. This always allows end users – as a rule linguists – to access the wide range of language data packed within INEL corpora using the set of linguistic analysis tools delivered by EXMARaLDA.

To reach a broader audience beyond those with linguistic expertise the project also aims to create engaging visual materials. The two main components of INEL visual interfaces both base their backend on the Elasticsearch engine: the Tsakorpus platform<sup>40</sup> repacks each INEL corpus as a web-based resource with a search interface that allows creating search queries through multiple layers of corpus annotations, and Kibana dashboards that, among other technical use cases within project workflows, are ideally suited for interactive visualisation of textual corpus data. At the moment, INEL offers Kibana-based digital versions of the Kuzmina archive<sup>41</sup> and the Dulson catalogue<sup>42</sup> that contain important manuscripts on Selkup and Nganasan languages (an extensive overview can be found in Lehmberg 2022).

Other visual resources include a map comprising all geographical locations where the language data that eventually ended up being included in the corpora were first collected. One more map is a "sequel" of the Kuzmina catalogue in Kibana and is a digitised version of a handdrawn map of the Taz River region from the manuscript collection (Lazarenko and Riaposov 2022).

<sup>&</sup>lt;sup>39</sup> www.exmaralda.org.

<sup>&</sup>lt;sup>40</sup> https://github.com/timarkh/tsakorpus/.

<sup>&</sup>lt;sup>41</sup> https://inel.corpora.uni-hamburg.de/portal/kuzmina/.

<sup>&</sup>lt;sup>42</sup> https://inel.corpora.uni-hamburg.de/portal/dulson/.

# 1.1. THE INEL PORTAL

On the INEL portal<sup>43</sup> one may browse through resources developed in the project, including but not limited to those described in the prior section. For example, the extensive and ever-growing INEL bibliography comprising publications of various kinds on minority languages of Siberia (Wagner-Nagy and Arkhipov 2019) could be of particular interest to language researchers. To date, the main operating language of the portal is German, as only the help section has been translated into English and Russian. This state of affairs sadly puts a limit on the reach the portal can have; given the precarious nature of the languages the project is concerned with, our goal guiding the current and future development efforts is to grow the project's presence on the web to be more accessible to audiences wider than German-speaking professional linguists. That, of course, includes the speaker communities, from which the language data came in the first place.

## 2. LANGUAGE COMMUNITIES AND LANGUAGE VISIBILITY

As stated, the languages INEL works with are extremely vulnerable in terms of their vitality. One of the four languages the project has published a corpus for – Kamas – has been extinct for over 30 years now with the death of its last speaker in 1989. The other three are in extreme danger: the number of speakers ranges from a little more than 5,000 for Evenki to ca. 1,000 for Dolgan, and a few dozen in the case of Selkup (Brykina, Orlova, and Wagner-Nagy 2021; Däbritz and Gusev 2021; Däbritz, Kudryakova, and Stapert 2022). The situation with Nenets and Nganasan unfortunately is not better – the number of speakers is rapidly declining. According to the catalogue of endangered languages, Nganasan, Forest Nenets, and Tundra Nenets are all endangered with Nganasan having extremely few native speakers tend to shift towards Russian already at a young age (Forest Nenets 2022; Nganasan 2022; Tundra Nenets 2022).

<sup>&</sup>lt;sup>43</sup> https://inel.corpora.uni-hamburg.de/portal/.

Since language endangerment criteria include, among others, the total number of speakers, looking at this criterion alone one can say that all these languages are, undoubtfully, ranked as at least severely endangered (Brenzinger et al. 2003: 8). Language communities are defined as a sum of speakers of the endangered languages (Khait, Lukschy, and Seyfeddinipur 2022: 540). Sadly, even in the digital era members of language communities do not always have easy access to the resources that linguists have gathered (naturally, from the communities themselves) and archived in order to preserve those languages. The traditionally stated goals of language preservation projects - recording and transcription of target language materials, preparation of grammar sketches, efforts to digitise older archives, and so on - have an understandable tendency to pivot towards the scholarly community as their primary target audience, the needs and wishes of which do not necessarily align well with those of the speaker communities of respective languages. Moreover, the linguistic research community could often be "too focused" on the grammatical properties of the gathered data rather than the contents of it which hold precious knowledge about local customs. The latter is indeed valuable in terms of preserving the endangered language not only as a collection of linguistic information but as a complex cultural phenomenon (Holton 2012: 106).

As a result, the speakers' engagement levels suffer. Traditionally the speakers of endangered languages would be offered the parts of language consultants and/or local folklore enthusiasts striving to aid professionals in their documentation efforts - the parts which, while admirably noble and, in cases of last speakers of moribund languages, bitterly tragic, could hardly be called engrossingly engaging for wider audiences. According to Khait, Lukschy, and Seyfeddinipur (2022: 540–541), until recently, it had been a common practice to give out the collected data to the communities saved on various physical storage devices with only a limited lifespan such as tapes, CDs, or even USB sticks, which did not speak for sustainable handling of the linguistic and cultural heritage of the communities. However, even with the shift to digital language archives and similar initiatives that advocate for data sustainability, access to those resources by the communities does not necessarily get easier. It has been often pointed out that the target audience of such data collection is quite a small one, which is caused by

two main factors: the interface language of the said resources is often English, and the way the data is presented is overly complicated for non-professional use, be it the way the data is structured, or the necessity to master specific tools to access and browse it (Khait, Lukschy, and Seyfeddinipur 2022: 540–541). This limits the target group of the users to the language specialists and/or people literate in English and inevitably leaves at least some if not all community speakers behind, keeping the language materials practically out of their reach.

The Internet - "the great equalizer" - fails such dubbing when it comes to languages: speakers of English are more equal than others. In the field of natural language processing any language outside of about 20 most widely researched ones runs a risk of being called "low-resource" (Cieri et al. 2016: 4543-4544; Magueresse, Carles, and Heetderks 2020: 1). It is the sad truth about many (yet living) languages of the world that they are practically invisible on the Internet, verging on effectively being "no-resource" with prospects growing dim. The Internet of endangered languages offers its scarce users to navigate through mounds of lingua franca content in order to arrive at that one target language page. Naturally, there are language enthusiasts and activists within the communities who try to keep their mother tongues alive; nevertheless, given that indigenous Siberian communities are quite small, building community language resources from linguistic corpora could be of great help to support them. Also, it would be in a way a "tribute" to the native speakers who help to collect and pre-process the raw field trip data. Holton (2012: 108-109) provides examples of successful use cases of using ANLA to create language resources that help to maintain or even revive interest in Alaska indigenous languages, such as Eyak. Schwiertz (2012) talks about ways in which a dedicated DoBeS archive web portal addresses different user groups from linguistic scholars to community speakers and the general public. After all, it is only logical that INEL should attempt to facilitate access to the released corpora to the indigenous speaker communities.

To sum up, it has been determined that there are several obstacles preventing community speakers from getting their hands on resources in their mother tongue. These are shared between several marginalised and endangered language communities (an extensive overview can be found in Khait, Lukschy, and Seyfeddinipur 2022: 541–542):

- poor technical infrastructure in the areas where the communities live - speakers might have only limited access to the internet, so one cannot expect even stable mobile Internet connection as a status quo in those areas;
- not every community speaker has a high level of digital literacy and this can hinder access to the data even more, especially when it comes to the linguistic resources that require specific software, or even a registration procedure/authorisation with an institutional ID in order to obtain the materials; moreover, many people prefer to use mobile devices rather than PCs and might, in general, have only limited online experience using only a small number of common applications and/or websites;
- it might be overly complicated to find those language archives on the web in the first place due to the way the website metadata is compiled.

That being said, inevitably few of the factors named above also apply to the main INEL product - linguistic corpora. First of all, it is compiled primarily in English regarding metadata and documentation. Although the project strives to always provide English, German, and Russian translation and glossing of the texts, it would be expected that users might want to familiarise themselves with the corpus documentation before actual browsing and the English language can become an obstacle. This adds up to the EXMARaLDA tools in English and makes the corpus data practically out of reach for those who are not skilled in English. Secondly, even though the Tsakorpus platform and Kibana dashboards are served on the web and can be accessed with mobile devices, they are not necessarily mobile-friendly in nature. In order to browse INEL corpora offline, one needs to use a desktop, preinstall and familiarise oneself with the EXMARaLDA software suite, and then download the said corpora which requires, in the best case, a non-metered, stable Internet connection, as the size of one corpus varies from several hundred megabytes to several gigabytes. The corpus documentation, metadata and transcriptions are overflowing with specific annotations and glossing. This is undoubtedly useful for linguists who would like to use these materials for their own research, however, for someone who would simply like to browse texts or media

in their native language, this can become overwhelming. Therefore, we raise such questions as "What can we do to give community speakers easy access to materials in their native languages?" and "How to make low-resourced languages more visible online?".

#### 3. SOLUTION

# 3.1. GENERAL CONCEPT

In order to address the challenges outlined in the previous section, we developed the following approach to present endangered language data online.

First, the materials we publish are openly accessible to anyone interested. Given that the data is publishable under GDPR (the EU General Data Protection Regulation) in the first place, we see no further motivation to restrict access in any way, shape, or form. The corpora are distributed under the CC BY-NC-SA 4.0 license.<sup>44</sup>

Second, as we effectively give the materials collected and archived by linguists in the years past back to the communities, it is paramount to present the data in a way that would be easily understandable by community members. By that, we mean (a) editing out excess linguistic notation – which the corpora are ripe with – from the community version of the materials, as it would cause more confusion than help; and (b) presenting the data in a language the community members would speak – in our case, it is Russian since all languages we have worked on to date are minority languages of Russia.

Third, in order to increase engagement levels across speaker communities, we pay much attention to the technical aspects of presentation: the web pages the materials are found on should be easily navigable from both mobile devices and desktops. More on that in the following sections.

Finally, community members are welcome not only to use the materials but to make contributions of their own as well. The relationship between linguists and speakers is not supposed to be a one-way street

<sup>&</sup>lt;sup>44</sup> https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.

after all, as some back and forth between these parties in our view would massively benefit both.

Therefore, the presentation of corpora data had to be redefined so that community members would get easy access to it and, above all, would have only positive experiences browsing it. The most straightforward option is creating a dedicated section on the website of the project that would serve as an entry point for the community speakers to familiarise themselves with the resources INEL has to offer. That is, developing a concept of the webpage, filling it with only relevant information in Russian and populating it with materials in the target languages.

#### 3.2. FROM INEL CORPORA TO INEL COMMUNITY

The most up-to-date version of the Community pages category is available through the INEL resources portal.<sup>45</sup> The first attempts at creating an INEL-community "spin-off" webpage date back to the beginning of 2021 to be presented during the INEL virtual "Day of Open Doors" (Open Day: Digital Presentation of Our Corpora 2021). Back then, there were two language pages created - Dolgan and Selkup. A landing page that offered a brief overview of the project tasks and goals and presented the project researchers served as an entry point to the Community pages. The structure of both Dolgan and Selkup pages followed the same principle: on top of the page, there was some introductory information about the language and links to the original INEL resources, followed by a simple table with a selection of the texts and basic metadata on the text genre and informant. INEL linguistic team defined which materials would populate the language pages. For the most part, linguists of the INEL team chose folklore and narrative texts, for example, fairy tales, legends, or personal stories on various cultural topics. Each table entry pointed to a separate HTML page with a text transcript and an embedded media source (if available). The transcripts were naturally not served as they would appear in the corpus - their structure was simplified and edited. Chosen EXMARaLDA

<sup>&</sup>lt;sup>45</sup> https://inel.corpora.uni-hamburg.de/portal/community/.

transcriptions were processed with XQuery scripts to extract relevant transcription tiers only – namely transcription in the target language and its translation to Russian – which were then transformed into HTML pages representing the text in a tabular form and embedding a time-aligned media source, if one is available, with the help of a simple JavaScript function.

Although the first version of INEL Community pages achieved the goal of presenting target language information alongside its Russian translation via a number of plain static HTML webpages within a dedicated section on the INEL portal, the pages themselves lacked user-friendliness and were hard to find using a search engine. With these challenges in mind, a new direction in the development of INEL Community pages took off from here on. The steps to undertake can be summarised as follows:

- new main page concept and development;
- redesigning existing language pages;
- creation of language pages for further project languages;
- concept and development of further materials.

When it comes to the concept, the core idea of it is user-friendliness. Given the technical limitations community members may be facing, the most logical approach is to keep the already existing structure yet make it intuitive and suitable for browsing with mobile gadgets. That is, the primary goal was to recreate the whole INEL community section starting from its main page. The next step was to generate a new community page for Nganasan following a request from our linguistic team. We maintain focus on the gradual upgrade of the older language pages – Selkup and Dolgan – so that they would fall more in line with the idea of user-friendliness. This is particularly important as it can potentially elevate the engagement of both community members and people who might be curious about endangered languages but not necessarily have the linguistic expertise to interact with the original corpus data with ease. Hence, it would draw more attention to INEL materials and the overall problem of language endangerment and cultural heritage.

# 3.2.1. MAIN PAGE

As stated previously, INEL Community pages exist within the main portal, however being entirely in Russian as the lingua franca of the modern indigenous Siberian communities, they are "on their own" within the portal structure. A user might well be confused about how to interact with the content and where to begin. Therefore, a main page is necessary. It dubs the content of the "regular" main page of the portal in Russian, yet not fully. Other than that, it also contains some of the general information from the project page on the University of Hamburg website. We decided to include the following information sections:

- general information about the project, its aims and goals;
- team presentation and contact data;
- links to the actual language pages;
- links to further project resources that might be of interest to the community speakers.

Technology-wise, we had a trade-off between maximum sustainability and user-friendliness. We abstained as much as possible from third-party software: while using it can result in an eye-pleasing website or web application and highly elevate user experience, its maintenance puts a lot of pressure on the team members in the long run. However, keeping just the HTML code convenient to read only from a desktop browser is also not the best choice, because even with CSS rules for better display on different gadget types it can be tricky to oversee all the ways how and with which means the end users interact with the data. Striving towards improved mobile user experience with the Community pages, we set the focus on the "mobile-first" design and development strategy and chose to use the Bootstrap frontend framework.<sup>46</sup> It allows the creation of mobile-first responsive websites that comply with the newest standard of web development, it is supported by most of the popular desktop and mobile browsers, and is highly customisable.

As for filling the main page with the information mentioned above, instead of providing everything as plain text section by section, we

<sup>&</sup>lt;sup>46</sup> https://getbootstrap.com/.

grouped these sections based on the content and wrapped them into visual blocks:

- проект INEL ("The INEL project") this block contains the first two information sections, namely a general short overview of the project, team presentation, and contact data;
- коллекция текстов INEL ("INEL text collection") links to the language pages themselves, initially Dolgan and Selkup, currently Nganasan as well; a more detailed overview can be found in the "Language Pages" subchapter;
- аудиовизуальные ресурсы ("Sound and visual resources") this section is currently in progress and by now is populated with links to geographical materials (see "The INEL Project" chapter) and in future also various interactive resources; more on that can be found in the "Further Resources" subchapter.

The sections are embedded into collapsible objects that allow page visitors to interact with the blocks by opening only relevant sections without having to scroll through all the content. Inside each collapsible element, one finds images with prompting titles, by clicking on them a text wrapped into modal objects appears, except for the language pages – clicking on them will redirect to the pages themselves.

This mode of presentation creates a compact and intuitive overview of the page components, especially on mobile gadgets, and is more visually engaging, prompting the user to explore the page further.

# 3.2.2. LANGUAGE PAGES

Initial Dolgan and Selkup language pages contained an introductory text about the language and the corpus as well as an HTML table containing the following information extracted from the corpus metadata file (COMA file):

- название ("Title") text titles with hyperlinks pointing to the webpages with the transcription;
- дата записи ("Recording date") the year when the respective recording was collected;

- место записи ("Recording place") the location where the respective recording was collected;
- диалект ("Dialect") a dialect of the respective speaker;
- рассказчик ("Speaker") each speaker in the corpus has a unique speaker code;
- звук ("Sound") information on whether the respective transcription/corpus entry contains audio;
- код текста ("Text code") transcription codes from the original COMA file.

Similarly, there are genre-based tables for the Nganasan language page, however, they are organised in a slightly different manner: instead of a speaker code from the COMA file, we see full names of the contributors, there are no dialect and text code columns, the sound column provides information on the availability of not only audio but also video sources.

Each page splits the total volume of available texts into several tables based on genres, among which one may find personal narratives, traditional songs, and legends. Again, unlike the COMA file, where one has to filter the data to find interesting resources, community page visitors can start browsing materials of any specific genre right away.

Each text has its dedicated page. As mentioned earlier, each page contains only the information that is relevant for a non-linguist – text in the original language, its translation in Russian, and, where available, the original media source.

Similar to the main page, the original language pages contained no corresponding stylesheets, therefore it was close to impossible to comfortably browse them from mobile browsers, given the fact that plain HTML tables cannot be rendered with correct scaling for mobile devices. The no-scaling-on-mobile problem affected texts on the transcription pages as well. Although the latter pages contained basic CSS styles, those were rather addressed to the font choice than to the responsive page design. After the language pages were partially rebuilt using Bootstrap plugins to gain a more interactive design, both text collection and actual transcription tables received custom CSS rules to adapt them to viewports of various sizes. The Nganasan texts were the first ones to receive "fully" upgraded language and text pages, the updates of Selkup and Dolgan ones are currently underway. Nevertheless, we do not plan to stop after this update: our goal is to continue to make the pages more and more user-friendly and responsive based on the feedback that we receive from the community and external website visitors as well as based on our internal UI/UX audits. Furthermore, there are two more tasks to be tackled. The first one is better integration of media sources into the text pages – this is especially important for the Nganasan community page, as there are multiple video sources – including improved audio synchronisation and providing subtitles for the videos. Another task is better search engine optimisation of the pages for their increased visibility on the web.

## 3.2.3. FURTHER RESOURCES

Anticipating what potential visitors of the Community pages might want to see apart from the relatively simple text collection, we attempt to provide them with several resources that are more interactive. In the first place, this is oriented toward the younger generation and is an attempt to further enhance the user experience with the INEL materials. Moreover, it will enrich the already existing visual resources with more mobile-friendly materials. This section of the INEL Community pages is at the moment in development, therefore we are describing the most up-to-date concept that is still subject to change. These materials will populate the section Audiovisual resources. By now, one can find there a short summary of all the INEL geographical data that is available at present. There will be at least two subsections available.

The first one is the so-called visual dictionaries. Inspired by other visual dictionaries out there (e.g., Visual Dictionary Online 2022; BABADADA 2022), we intend to create a simple visual dictionary for Dolgan, Selkup, and Nganasan. In essence, it should be a gallery of culturally relevant pictures, such as photos taken during field trips, with labelled objects, whereby hovering over, one can see what this object is called in the target language, its translation to Russian and point to text and audio samples containing that word from the corpus. We expect that these dictionaries will become another visual entry point to the INEL resources and motivate people to dive into the textual materials through the visual ones. Last, but not least, it might become a helpful educational tool to refresh or learn the vocabulary of the native language, which we consider to be of great importance in case of endangered and shifting languages.

Another concept with a similar mission of visual data presentation is a tag cloud or word cloud. Here, the focus is not on linking a word with its visual representation, but rather on the graphic representation of word lists extracted from the corpora. More specifically, for a test run, we extracted ca. 50 most frequent nouns from the Selkup corpus and fed them to Kibana which resulted in a simple tag cloud. Further, we intend to create such tag clouds for other languages and parts of speech and link them to the text data. All in all, it would be another way to present the corpus data visually and provide an overview of topics covered in the texts.

#### CONCLUSION

We have presented our approach to bridging the gap between the community of linguistic experts who aim at documenting endangered languages, and the speakers of those languages who often have little to no access to said documentation.

To sum up, creating language archives and corpora is a great way to preserve the valuable linguistic and cultural data of endangered languages. However, it is crucial to pay heed to the speaker communities as well. There is something wrong with situations where the data provided by the community ends up collecting dust in archives and repositories, which require an institutional account or completion of some complex procedure in order to be accessed. Another obstacle is the structure of such datasets, which is often oriented at researchers and therefore not comprehensible enough without professional expertise; moreover, knowledge of the English language is almost always a musthave to be able to interact with the resources. All this creates a certain paradox – even though an endangered language might look "backed up" on paper as it is well described, it remains endangered in the field – and the poor accessibility of resources runs counter to the idea of preserving cultural heritage, of which language itself is a big part. Therefore, we propose that linguists take into account the possibility of supplying online-based user-oriented materials alongside resources intended for professional use. The online part is of particular importance here, as it considerably increases the outreach and allows community members to take a more active part in the process. However, getting community members to rally around online resources designed to support and promote cultural and linguistic heritage is not that easy, and approaches to achieve that need further coverage.

In this paper we have presented our way of doing it – starting from a general concept to more specific ideas. Naturally, as it is only a work in progress, the INEL Community pages will undergo further development and modification based on the users' feedback.

#### **REFERENCES LIST**

- "BABADADA". 2022. BABADADA GmbH. Available at: https://www.babadada. com/ (accessed 10 November 2022).
- Brenzinger, Matthias, Arienne M. Dwyer, Tjeerd de Graaf, Colette Grinevald, Michael Krauss, Osahito Miyaoka, Nicholas Ostler, Osamu Sakiyama, María E. Villalón, Akira Y. Yamamoto, and Ofelia Zepeda. 2003. "Language Vitality and Endangerment". Document adopted by the International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages, Paris, 10–12 March. Paris: UNESCO.
- Brykina, Maria, Svetlana Orlova, and Beáta Wagner-Nagy. 2021. "INEL Selkup Corpus". Version 2.0. Publication date: 31 December 2021. Available at: https://hdl.handle.net/11022/0000-0007-F4D9-1 (accessed 2 November 2022). Archived at the University of Hamburg. In The INEL Corpora of Indigenous Northern Eurasian Languages. Available at: https://hdl.handle. net/11022/0000-0007-F45A-1 (accessed 2 November 2022).
- Cieri, Christopher, Mike Maxwell, Stephanie Strassel, and Jennifer Tracey. 2016. "Selection Criteria for Low Resource Language Programs". In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, eds. Portorož, Slovenia: European Language Resources Association, 4543–4549.
- Däbritz, Chris Lasse, Nina Kudryakova, and Eugénie Stapert. 2022. "INEL Dolgan Corpus". Version 2.0. Publication date: 30 November 2022. Avail-

able at: https://hdl.handle.net/11022/0000-0007-F9A7-4 (accessed 4 December 2022). Archived at the University of Hamburg. In The INEL Corpora of Indigenous Northern Eurasian Languages. Available at: https://hdl.handle.net/11022/0000-0007-F45A-1 (accessed 4 December 2022).

- Däbritz, Chris Lasse and Valentin Gusev. 2021. "INEL Evenki Corpus". Version 2.0. Publication date: 31 December 2021. Available at: https://hdl.handle.net/11022/0000-0007-F43C-3 (accessed 30 October 2022). Archived at the University of Hamburg. In *The INEL Corpora of Indigenous Northern Eurasian Languages*. Available at: https://hdl.handle.net/11022/0000-0007-F45A-1 (accessed 30 October 2022).
- "Forest Nenets". 2022. Catalogue of Endangered Languages. University of Hawai'i at Manoa. Available at: https://www.endangeredlanguages.com/lang/3393 (accessed 30 October 2022).
- Holton, Gary. 2012. "Language Archives: They're not just for Linguists any more". In Potentials of Language Documentation: Methods, Analyses, and Utilization. Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts, and Paul Trilsbeek, eds. Honolulu: University of Hawai'i Press, 105–110.
- Khait, Ilya, Leonore Lukschy, and Mandana Seyfeddinipur. 2022. "Linguistic Archives and Language Communities Questionnaire: Establishing (Re-) use Criteria". The Electronic Library 40 (5): 539–551. Available at: https:// doi.org/10.1108/EL-01-2022-0012 (accessed 1 December 2022).
- Lazarenko, Elena and Aleksandr Riaposov. 2022. "How to Digitize Completely: Interactive Geovizualization of a Sketch Map from the Kuzmina Archive". In Proceedings of the Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-Resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference. Atul Kr. Ojha, Sina Ahmadi, Chao-Hong Liu, and John P. McCrae, eds. Marseille, France: European Language Resources Association, 22–27. Available at: https://aclanthology.org/2022.eurali-1.4/ (accessed 20 October 2022).
- Lehmberg, Timm. 2022. "Not Sustainable but Beautiful? Some Steps towards Visual Access to Multidimensional Data Collections". Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology 7 (September): 71–79. Available at: https://doi.org/10.14232/ wpcl.2022.7.4 (accessed 30 October 2022).
- Magueresse, Alexandre, Vincent Carles, and Evan Heetderks. 2020. "Low-Resource Languages: A Review of Past Work and Future Challenges". *ArXiv.* arXiv/2006.07264 [cs.CL]. Available at: https://doi.org/10.48550/ arXiv.2006.07264 (accessed 30 October 2022).
- "Nganasan". 2022. Catalogue of Endangered Languages. University of Hawai'i at Manoa. Available at: https://www.endangeredlanguages.com/lang /2893/ (accessed 30 October 2022).
- "Open Day: Digital Presentation of Our Corpora". 2021. INEL. University of Hamburg. Available at: https://www.slm.uni-hamburg.de/inel/aktivitaeten/aktuelles/2021-2-23-open-day.html (accessed 30 October 2022).

- Schmidt, Thomas and Kai Wörner. 2014. "EXMARaLDA". In *Handbook on Corpus Phonology*. Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, eds. Oxford, UK: Oxford University Press, 402–419.
- Schwiertz, Gabriele. 2012. "Online Presentation and Accessibility of Endangered Languages Data: The General Portal to the DoBeS Archive". In Potentials of Language Documentation: Methods, Analyses, and Utilization. Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts, and Paul Trilsbeek, eds. Honolulu: University of Hawai'i Press, 126–128.
- "Tundra Nenets". 2022. Catalogue of Endangered Languages. University of Hawai'i at Manoa. Available at: https://www.endangeredlanguages.com/lang/2062 (accessed 30 October 2022).
- "Visual Dictionary Online". 2022. QA International. Available at: http://www.visualdictionaryonline.com/index.php/ (accessed 10 November 2022).
- Wagner-Nagy, Beáta and Alexandre Arkhipov. 2019. INEL Bibliographie (Version 1.0) [Dataset]. Available at: http://doi.org/10.25592/uhhfdm.731 (accessed 20 October 2022).

# ENHANCING ACCESS TO DIGITISED GLAGOLITIC MATERIAL OF ZADAR COUNTY THROUGH GLAGOLAB PORTAL

# LAURA GRZUNOV

University of Zadar, Department of Information Sciences Zadar, Croatia Igrzunov@unizd.hr

# MARTA IVANOVIĆ

University of Zadar, Office for Publishing Activities Zadar, Croatia mivanovi1@unizd.hr

# MARIJANA TOMIĆ

University of Zadar, Centre for Research in Glagolitism / Department of Information Sciences Zadar, Croatia mtomic@unizd.hr

Today, cultural heritage (CH) collections are being increasingly digitised and becoming available to everyone via the web. In order to become more searchable, more visible and available for various research in such a virtual environment, the search box alone, which is the most usual search option, does not suffice. Scientific Centre for Research in Glagolitism of the University of Zadar is conducting a range of interdisciplinary, cross-institutional projects with activities that include the digitisation of Glagolitic manuscripts of Zadar County and its fragments, digitisation of watermarks found in those manuscripts and designing and building a database of Glagolitic manuscripts, fragments and watermarks. The GlagoLab portal was built as part of the Centre and is intended for creating a common repository for research of the Glagolitics of Zadar County. After digitising and publishing Glagolitic manuscripts, fragments, and watermarks, research and access possibilities need to be explored. The focus of this paper is on supporting access to digital collections of Zadar Glagolitics, both for scholars and for the general public. To search digital collections of Zadar Glagolitics, it is expected that scholars use search boxes, but, at the same time, they are browsing catalogues to discover material of interest. This paper will analyse and discuss browsing strategies with the aim of developing an interface with search and browsing options that will facilitate efficient exploration of Zadar Glagolitics digital collections.

Keywords: cultural heritage, Glagolitic, searching, browsing, Glago-Lab portal

#### INTRODUCTION

This paper analyses theoretical assumptions about interfaces for searching and browsing digital cultural heritage collections. It emphasises the need to move from a search box to an interface that will facilitate the efficient exploration of valuable cultural heritage.

The first chapter, "From Search Box to More Liberal Interfaces?", introduces the theoretical background to the topic of accessing, searching, browsing and using digital cultural heritage collections in different ways. This chapter explains the reasons for switching from the search box to an interface that will satisfy different types of users and their diverse needs.

On the basis of the topic set in the theoretical settings, the aim, phases, and methods used for the purposes of the research described in this work are presented in the second chapter, titled "Research Aim, Phases, and Methodology".

The third chapter, "A Shift from Searching to Browsing", explains the shift from search boxes to exploratory search interfaces and to interfaces that enable browsing of digital cultural heritage collections (rich-prospect browsing and generous interfaces). In order to make it clearer to the readers what each interface refers to, this chapter analyses examples of various projects that provide such interfaces.

In the fourth chapter, "GlagoLab portal", an analysis of the aforementioned portal developed as part of the Centre for Research in Glagolitism at the University of Zadar is presented. Through the analysis, an effort was made to see the possibilities regarding display options and tools for working with Glagolitic manuscripts.

The "Conclusion" summarises the results of the analysed literature, projects and final considerations regarding the possibility of searching and browsing the digital Glagolitic heritage on the GlagoLab portal.

#### 1. FROM SEARCH BOX TO MORE LIBERAL INTERFACES

In the last few decades, digitisation has made it possible to make a wealth of cultural material available on the Internet in digital form. Many digital collections of cultural objects are accessible to everyone via the web, both to scientists who are researching them, and also to the general public. Because digital collections serve diverse audiences with different needs, more thought needs to be given to ways of accessing, searching, browsing, and using these collections. It has been shown that galleries, libraries, archives, and museums (GLAM institutions) lose more than half of their users after they enter the first page (Hall and Walsh 2021: 261). The reason for this is the search interfaces that do not show the abundance and wealth of materials and objects that the collection possesses. The search interface is the place where searchers interact with the search system and it should be able to help users resolve different information problems in both their working and living environments, as well as support users in finishing the entire work task or achieving their information goals, and not only support the search aspect (Liu et al. 2021: 3-4). "Search interface as embodied by the text box and keyword has framed our understanding of what the Web is" (Wilson et al. 2010: 6). Search interfaces, i.e., an empty text field, require the user to come up with a query without often knowing what they can type and what they are looking for. Such interfaces expect the user to be familiar with the digital collection and to know the keywords and terms they want to search by. For this reason, such a search is quite problematic for users who encounter a digital collection for the first time and do not know the material it possesses or the terminology that can be used for its search. Digital collections of cultural objects "typically offer keyword-based search interfaces, which are well-suited for expert users but which do

not support non-expert users, who are often unfamiliar with the collections and struggle to formulate appropriate queries" (Hall 2014).

Users visit the websites of GLAM institutions and their digital collections for many reasons. Therefore, when building such collections, the different needs of users, different profiles of people and the reason for entering the website should be taken into account. The level of knowledge and expertise about specific cultural objects varies significantly from user to user, and for this reason, when building digital collections, different ways of searching and browsing should be considered. Supporting the diversity of potential users and their requirements is quite difficult precisely because there is a difference in knowledge of a certain digital cultural heritage collection. Users who encounter a search field for the first time do not know if they will get any results and most often, they don't even know what they should type in that field. On the other hand, experts who are already familiar with the digital collection or who use materials from the digital collection as a corpus of research know exactly what they are looking for and what they need to enter into the interface in order to get results. The very high bounce rates of over 60% experienced by GLAM institutions' online sites indicate that there is a significant proportion of potential visitors for whom current digital collection search capabilities do not work (Hall and Walsh 2021: 261). It was previously stated that users often give up searching immediately after entering the first page. The reason for giving up can be precisely the lack of knowledge of terminology, keywords, titles, and items that can be searched in the collection. Because of that the need for richer interfaces that reveal the richness of the collection with the possibility of browsing is increasingly emphasised. Today, most often when searching for digital cultural heritage collections, we come across a search box that allows us to search by title, subject, and keywords. Searching by keywords turned out to be insufficient and poor. Such interfaces fail to be liberal in sharing and, as Whitelaw (2015) emphasises, instead of throwing open the doors, its greeting is "Yes, what?". To use such interfaces, the user must know the collection and features of the material being digitised in order to be able to search successfully, while those who do not know enough about the collection and its features will not find everything they are looking for.

# 2. RESEARCH AIM, PHASES, AND METHODOLOGY

The aim of this work was to investigate theoretical approaches to enhancing access to cultural collections in order to analyse and propose searching and browsing strategies for digital collections of Zadar Glagolitics.

The research consists of three phases:

- 1<sup>st</sup> phase to present the theoretical framework about providing and enhancing access to cultural collections, with the aim of researching different ways of searching and viewing digital collections that could be applied on the GlagoLab portal where digital reproductions of Glagolitic manuscripts and watermarks on paper are available;
- 2<sup>nd</sup> phase involves surveying users (experts and the general public) about their preferences and needs when searching and browsing digital collections;
- 3<sup>rd</sup> phase application of research results from the preceeding phases and implementation of new collection search strategies on the GlagoLab portal.

For the purposes of this paper, the first phase of the research will be presented. The first phase includes various search and browsing options which were analysed in order to see which of those are applicable on the GlagoLab portal and whether they correspond to the settings of searching and browsing options of digital cultural heritage collections that have been proposed in the literature. For the purposes of conducting the first phase, the interfaces of digital collections were analysed through:

- literature review;
- analysis of the interface of various projects on The European Association for Digital Humanities (EADH) association's website, which "brings together and represents the Digital Humanities in Europe across the entire spectrum of disciplines that research, develop, and apply Digital Humanities methods and technology" (EADH n.d.). On the website of the EADH association there is a list of projects undertaken during the last five years that that make a meaningful contribution to digital humanities in Europe. The list of

projects will be used for the 1<sup>st</sup> phase, in which different ways of searching and viewing digital collections are explored;

 analysis of the search and browsing interface of the GlagoLab portal to see which search and browsing options are currently available and whether they correspond to the settings of the generous interface.

## 3. A SHIFT FROM SEARCHING TO BROWSING

It is clear that search interfaces are not a sufficient option, especially when the different users' needs are taken into account. Search interfaces require querying, discourage exploration, and withhold more than they provide (Whitelaw 2015). It is necessary to develop interfaces that will allow the user to find sources in digital collections accidentally and in a targeted manner.

Whitelaw (2015) emphasises that the model of interface and interaction, in the field of digital humanities, should encourage exploration and interpretation, not just searching and finding information. So, the main question is: Why don't we provide other alternatives in addition to the search option, and what alternatives are there?

The search interface is a barrier for many users that have no previous experience with the topic and therefore need more help and information when they start browsing the collection. It is for such users that White and Roth (2009) propose "exploratory search systems".

# 3.1. EXPLORATORY SEARCH INTERFACES

Hall and Walsh (2021) wrote about exploratory search interfaces in their work *Exploring Digital Cultural Heritage through Browsing*. They emphasise that such interfaces "are generally designed to provide the user with guidance as to which keywords will produce search results and to help them narrow down their search, using features such as query suggestion and search facets" (Hall and Walsh 2021: 262). The authors clarify that, although such interfaces give users a form of instructions, this way of searching still does not allow an overview of the collection as a whole. This would mean that the user must know what he is looking for in the collection and that the collection was created for the needs of those who have some kind of goal.

The most common example of exploratory search is the "faceted search interface". According to Russell-Rose and Tate (2013: 167), "a key principle of faceted search is to minimise the likelihood of zero results by guiding users toward productive navigational choices. In practice, this means displaying only currently available facet values (i.e., those that apply to the current navigational context), and eliminating those that would lead to dead ends". In the previously mentioned work by Hall and Walsh (2021), in which the authors are focused on the presentation of additional options for searching and browsing items, they also talk about the faceted search interface. They point out that the advantage of the faceted search interface is that "instead of being required to enter a search term, the user can select a value from the facet list and see results for that value. The exact facets used depend on the available meta-data, but commonly available facets include dates, locations, categories, materials and techniques" (Hall and Walsh 2021: 264). Faceted search enables non-expert users to learn what search terms will lead to results.

Although there are various advantages of faceted search, the problem is that this type of search is still focused only on search. To delve further into the question posed in this paper, when talking about possible alternatives to searching, the literature also mentions "browsing".

#### 3.2. BROWSING

Browsing the shelves, which is enabled for users in the physical spaces of GLAM institutions, should now move to a digital environment so that users can become aware of the material that exists, because what cannot be seen, they cannot even look for (Hibberd 2014).

Current GLAM institutions' interfaces allow for targeted searches rather than discovery. Hall and Mark (2021: 262) state that a browsingbased interface is one that allows the user to interact with the collection without having to explicitly enter a search keyword. The browsing option is crucial in digital collections, especially for less expert users with undefined goals for whom an empty search field is a big barrier. These types of users are also more likely to visit the collection in their leisure time rather than in a work context and thus tend to follow more exploratory behaviour (Mayr et al. 2016). Whitelaw (2015) defines browsing as an "iterative process that entails scanning or glimpsing a field of potential resources, and selecting or sampling items for further investigation and evaluation".

When developing browsing interfaces, there are 2 important requirements (Hall and Mark 2021: 265):

- they need to provide an initial overview of the collection;
- through the browsing and visualisation interface, they must support the user in exploring the collection and gradually building up a more detailed understanding of the content.

According to Hall and Mark (2021: 265), frequent examples of browsing digital collections in GLAM institutions are:

- a manually curated digital exhibition that provides an overview of the collection and very detailed information about the selected set of objects;
- browsing through the classification system (UDK, DDC).

Given the possibilities of digital technology today, in the context of browsing digital collections, the focus should be on more informative, scalable browsing interfaces that support the needs of those users who encounter the collection for the first time (Hall and Mark 2021: 265).

Such interfaces are referred to in the literature as "rich prospect browsing" and "generous interfaces". The two terms were developed independently, but essentially describe the same basic idea of providing an interface that does not require a priori expertise in either the interface or the collection in order to use the interface successfully. Although they differ in some respects, they both place a strong focus on providing the user with an overall overview of the collection, and on enabling exploration of said collection without the need to enter a search query (Hall and Mark 2021: 265).

# 3.2.1. RICH PROSPECT BROWSING

Rich prospect interfaces represent an entire collection at once, where the user has a meaningful representation of every single item in the collection. Interactive control and detailed access to elements should help users understand, on a visual basis, what is available in a collection so they would be able to explore the collection (Glinka et al. 2017). The guiding principle of rich prospect browsing is "first overview, zoom and filter, then details on request" (Hall and Mark 2021: 265). Users can manipulate items by filtering, zooming in and out, marking and linking to see more detail. Being able to see an entire collection allows the user to find or discover items and, what is more important, to understand the entire collection itself.

New York Public Library interface is one of the examples of rich prospect browsing (Figure 1).



Figure 1: New York Public Library Interface.

The home screen shows a grid with a small thumbnail of each item. The images on the home screen are organised by time, but can also be arranged by genre, collection or colour. Because each thumbnail is only a few pixels in size, the big question is whether this interface can serve as a meaningful representation of the individual items, because it's very difficult to make out anything about the items from a lot of tiny thumbnails.

By selecting a photo from the grid, the user gets a metadata description of the photo and the collection to which a certain object belongs within the library, which is quite useful. The user can, if interested, browse the contents of that collection further. There is also a timeline display in which the most important dates related to the object the user is viewing are listed.

#### 3.2.2. GENEROUS INTERFACES

A generous interface is an interface that should allow initial insight and overview of the collection. Unlike with rich prospect browsing, the assumption is that the home screen shows a sample drawn just from the collection and not all the items. The sample provides a starting point and clues to help the user explore the collection.

Generosity in generous interfaces, in contrast to the search slot implemented as a default starting point in many digital interfaces, is evident through five principles (Windhager et al. 2019):

- show first, don't ask;
- provide rich overviews;
- provide samples;
- provide context;
- share high-quality primary content.

Australian Prints and Printmaking is one of the examples of generous interface that is often mentioned in literature (Figure 2). It is a typical combination of rich prospect browsing and a generous interface because it offers different options to view items from collections such as the timeline, word clouds, and grid browsing. The free online collection contains over 22,000 images that are based on the Australian print collection at the National Gallery of Australia. The databases can be searched by artist, subject or print techniques.

If users already know the object of their interest, there is an interface with a simple and advanced search. But if they do not know the exact search word (artist, work, exhibition, etc.) or are just interested in exploration and discovery of the collection then there is an option to explore five unique interfaces: Subject Explorer, Timeline, Works and Networks, Decade Summary, and All Artists. Visualisation of large cultural heritage datasets is achieved through implementation of timelines, word clouds and grid browsing.



Figure 2: Australian Prints and Printmaking Interface.

Generous interfaces use multiple representations to reveal the complexity and diversity of cultural collections, such as visualisations. The most common visualisation methods include:

- timelines;
- spatial (map) displays;
- network diagrams;
- word clouds.

Timelines are the simplest solution for mapping time to space in a linear, one-dimensional fashion. They commonly visualise events as marks along a line (Windhager et al. 2019).

"History 503: Ancient Rome" project of creating a historical timeline was created by using TimelineJS, an open-source tool that enables anyone to build visually rich, interactive timelines (Figure 3). This student project was created through a collaboration between a history professor and a digital humanities librarian on an upper-division Roman History course, in order to prepare students to create a high-stakes DH project of their own. The goals of this collaborative project were for the students to: experience working collaboratively and to help them decide, in a low-stakes environment (on the timeline), whether this is the platform they want to use for their final/solo project; then, to demonstrate that the events emphasised in the construction of a timeline impact their understanding of fundamental skills for chronological and historical thinking for Roman history, as well as recognise the possibilities of media-rich digital humanities ways of sharing Roman history content.



Figure 3: "History 503: Ancient Rome" project.

We can use spatial displays to visualise spatial metadata on a two-dimensional map. The advantage of this type of map is that the user can zoom out to see an overview of the collection and zoom back in to see individual items. The most significant limitation of maps is that they require items to have spatial metadata that need to be in a computerreadable form with the information of the exact location, so that the current web-based maps can handle spatial information.

"Icelandic Saga Map" project is an example of using a spatial display to visualise data (Figure 4). It is created as a resource to guide specialists and non-specialists around the *Íslendingasögur* texts ("Sagas of Icelanders") from a spatial perspective. All of the *Íslendingasögur* have been geo-referenced, as well as *Landnámabók* ("The Book of Settlements"). Travel diaries and accounts by 19<sup>th</sup> century saga enthusiasts are also being geo-referenced.



Figure 4: "Icelandic Saga Map" project.

Network diagrams are non-temporal visualisation methods, often combined with other methods, and allowing users to explore the proximities and distances of CH objects in relational or topological spaces (Windhager et al. 2019).



Figure 5: "Histograph" project.

"Histograph" is an effective interface to access collections of historical sources (collection of more than 3,000 images which represent the main events and actors in the history of European integration) and to discover links among and entities within them (Figure 5). It is an example of treating large multimedia collections as networks through a force-directed graph layout that places frequently co-occurring nodes closer together, while nodes that co-occur less often are placed farther apart. "Histograph" was developed in an interdisciplinary collaboration of computer scientists, historians, HCI researchers, and interface designers. To create the network, faces in the historical photos needed to be identified, a challenging task for machines alone. Humans and machines, therefore, share this work and the input from one improves the performance of the other.

Word (tag) clouds are a prominent method of visualisation and verbalisation to represent metadata aspects of the collection. Tags/ keywords can be derived from existing object classification/titles/ description or generated through crowdsourcing/computer-vision methods (Windhager et al. 2019).

# Victorian Illustrated Shakespeare Archive

Rare Visions...

Actilies Acters Angels Animalis Antychius of Spraces Antenia Arman Army Athur Andreyan Ae Addry Authycia Areas Bables Ghoreb Bars Backsta and Bares Bassinio Bats Beds Beheadings Batarius Birds Binn Blocks Indies Bees and Arross Bables Charses Cassing Cherubis and Bables Charses Chairs Cassing Cherubis and Bables Charses Chairs Charse Bassinio Pilo Charses Cassing Charses Charse Bassinio Pilo Charses Charses Chairs Charses Charses Pilo Charses Charses Chairs Charses Charses Pilo Charses Charses Charses Charses Conflict Candida Castard Charses Andrey Charses Charters Charses Charses Charses Charses Pilo Pilo Terms Charses Charses Dake of Charses Charses Charses Pilo Coorvays Drink Charses Charses Charses Fare Feather Filers Filorge Florida and Falanca Charses Dake of Charses Charses Charses Charses Pilong Founting Charses Charses Charses Marker King Charses Pilonge Harses Herath Inter Pilong Barses Founting Charses Charses Moon Buddeness Mats Charses Charses Moon Buddeness Mats Charses Charses Moon Buddeness Mats Charses Charses Pilong Harses Pilonge Charses Pilonge Harses Pilong Harses Pilonge Pilonge Charses Pilonge Charses Pilonge Harses Pilonge Charses Pilonge Charses Pilonge Harses Pilonge Charses Pilonge Harses Pilonge Harses Pilonge Pilonge Pilonge Pilonge Charses Pilonge Harses Pilonge Pilo





#### Gaze Therein...

Charles Knight (ed.) 👻 Kenny Meadows 👻 John Gilbert 👻 H.C. Selous 👻

John Dicks (pub.) \* John Tallis (pub.) \* John G. Murdoch (pub.) \* Heroines of Shakspeare \* Songs & Sonnets of Shakespeare \*

> Act I Act I Headers Act II Act II Headers Act III Act II Headers Act IV Act IV Headers Actes Act V Act V Headers Afrik Will Thill Ends Will A Midsummer Nights Derain Antery and Chepatra Acti A Visu (Le B Endstee Derain Attes) Editors Catiban Charles

Knight Cleopatra Comedy Coriola (Characte 877 topics Tay) Cressida Cymbeline

Exterior Ferdinand (The Tempest) Full Page

Illustrations H.C. Selous Hamlet's Ghost Hamlet (Play) Helena (All's Well That Ends Well) Hermione

Heroines of Shakspeare History lago Imogen

Interior Introductory Remarks / Notice Isabella John

Dicks John G Murdoch John Gilbert John Talls

Juliet Julius Caesar Katharina Kenny

Meadows King Claudius King Herry IV Part I King Henry IV Part I King Herry V (Characted) Xing Henry V (Plur) Xing Herry VII (Characted) Xing Herry VII (Plur) Xing Herry VII (Characted) Xing Herry VII (Plur) III King Juhn (Characted) Xing Juhn (Plur) Xing Lear (Characted) Xing Lear (Plur) Xing Khardl II (Characted) Xing Richard III (Plur) Xing Richard III (Characted) Xing Richard III (Plur) Xing Khard Lieotter Lewix Labour's Leas Machelli (Characted) Michaell Plur) Mahaleli Marc Antory Measure Fer Measure Miranda Mistersa Calcity Mach Ada About Nething Ofwia Ophia Chande Othendo (Characted) Othendo Plur) Plur Cherkencho Photographe

Figure 6: The Victorian Illustrated Shakespeare Archive.

One of the examples of using word clouds in digital humanities projects is "The Victorian Illustrated Shakespeare Archive", an online open access resource that contains over 3,000 illustrations taken from the four major editions of Shakespeare's Works in the Victorian period (Figure 6). Using Photoshop to isolate the illustrations, the author single-handedly scanned more than 3,000 illustrations from hard copies of the play collections, tagging each image. The archive allows users to use word clouds to search the illustrations by motif. Users can search by each of the four editions and access each individual play's digitised illustrations by type: "histories", "comedies", and "tragedies".

It is evident that more and more digital humanities projects use rich prospect browsing and generous interfaces to make CH collections visually attractive, interactive, and more accessible to a larger number of users.

#### 4. GLAGOLAB PORTAL

For the purposes of the 1<sup>st</sup> phase of the research, the GlagoLab portal was analysed in order to see the possibilities regarding display options and tools for working with Glagolitic manuscripts. Based on the theoretical settings and the analysis of various search and browsing options, an analysis of the GlagoLab portal was made to see which search and browsing options are currently available and whether they correspond to the settings of the generous interface.

GlagoLab is a portal of the Centre for Research in Glagolitism at the University of Zadar and it is also a digital laboratory for collaborative research and promotion of the Croatian Glagolitic script. Centre for Research in Glagolitism is a scientific research centre of the University of Zadar whose main goal is the "interdisciplinary study, interpretation and promotion of Croatian Glagolitic, especially its Zadar department, i.e. coordinating the activities of scientists focused on Glagolitic research and creating assumptions for their work" (GlagoLab n.d.).

The GlagoLab portal contains a catalogue of digitised Glagolitic manuscripts and watermarks found in these manuscripts. When entering the portal, a search box is available on the main page, which, according to the literature, may be a useful option for experienced scientists, but not so much for the general public who do not know what to search for and in what way. Figure 7 shows the main page of the portal and the search box option.



Figure 7: GlagoLab portal.

Via the title page, users can access both the Catalogue of manuscripts and the Catalogue of watermarks (Figure 8).



Figure 8: Catalogue of manuscripts and watermarks on GlagoLab portal.

By entering the Catalogue of manuscripts or the Catalogue of watermarks on the left side is the possibility of faceted search, which, as we have previously established theoretically, belongs to exploratory search interfaces. Users can browse according to several facets – for manuscripts according to the type of material, place, and institution, and watermarks according to place, motif parts, and institution. It is certainly necessary to survey the two types of users (the scholars and the general public) about whether this kind of viewing option is useful for them and what aspects would be useful for them. In the catalogues, there is also a search box that enables searching by metadata, which is quite difficult for someone who is not familiar with the collection and not even with the term "metadata" itself. The option of viewing according to the grid is also enabled (Figure 9).



Figure 9: Searching and browsing possibilities on the GlagoLab portal.

The option "grid" shows a thumbnail with each item in both collections (Figure 10). By clicking on any thumbnail, the user will get an insight into the description of the manuscript or watermark and will be able to access the digitised versions. This display is similar to the NYPL example and can be considered an example of rich prospect browsing.



Figure 10: Grid option on the GlagoLab portal.

In addition to catalogues, a thesaurus of watermarks is also available on the GlagoLab portal (Figure 11). As part of the "Written Heritage" project, a thesaurus of watermarks was translated, which was created for the needs of the "Bernstein – The Memory of Paper" EU project.<sup>47</sup> The thesaurus defines the terms that will be used for the definition of motifs found on the watermarks within the manuscript. Each of the terms from the thesaurus has its associated icon, according to which users can browse the catalogue of watermarks, and by clicking on a certain icon, a list of all watermarks with that motif is obtained, as well as a visual representation of the motif in the thesaurus itself (whether it is a main term or a subordinate term of the thesaurus).

<sup>&</sup>lt;sup>47</sup> "Bernstein – The Memory of Paper" project is available on this link: https://www. memoryofpaper.eu/BernsteinPortal/appl\_start.disp.


Figure 11: Thesaurus enhanced search on the GlagoLab portal.

According to the analysis and the display of all search and browsing options, it can be concluded that GlagoLab is missing segments of a more generous interface:

- home page should show a sample drawn from the collection that provides a starting point and clues to help the user explore the collection;
- the interface should display options upfront rather than requiring the user to write a query, as they may not know what they are looking for or what to include in a query;
- using more visualisations such as:
  - timelines the users would be able to make comparisons and connections between different aspects of manuscripts or watermarks such as year of production and places;
  - spatial diagrams the collection contains manuscripts and watermarks from different places, and spatial diagrams can be used to illustrate for example the potential paper trade routes;
  - word clouds with standardised metadata, it would be very easy to create word clouds and that could help users with initial research.

Further research, that will include users of the GlagoLab portal and their preferences, is needed. Therefore, the 2<sup>nd</sup> phase of this research is to survey users – the experts and the general public – about their preferences when searching and browsing digital collections, in order to see how the GlagoLab portal can become more generous.

#### 5. CONCLUSION

This paper has aimed to highlight the significance of considering the lifecycle and accessibility of digital cultural heritage, which increasingly depends on how cultural heritage reaches the public and its public accessibility. It is clear that users today expect an interface where they can explore, and not just search the collection. This is why there is an increasing emphasis on visual information retrieval, which should definitely offer a richer experience, interaction, and interpretation of the collection. Although the literature often emphasises the importance of more generous interfaces, many interfaces are still stuck only on search capabilities. This paper gives examples that can serve as inspiration for expanding the possibilities of searching and viewing cultural heritage collections in a virtual environment. The 1st phase of research focused on literature review, analysis of the interface of various existing projects in the field of digital humanities, and the analysis of the search and browsing interface of the GlagoLab portal. The GlagoLab portal, which contains digitised Glagolitic manuscripts and watermarks, was analysed to determine whether the features of generous interfaces are missing in this collection of valuable Zadar cultural heritage. The analysis showed that the portal contains some segments of exploratory search interfaces and provides the option of browsing within the catalogues of manuscripts and watermarks. However, there is certainly a lack of options to help non-expert users browse the collection. For this reason, it is necessary to conduct the 2<sup>nd</sup> and 3<sup>rd</sup> phases of research, which involve surveying users (experts and the general public) about their preferences for searching and browsing digital collections and further interface research regarding display options and tools for working with Glagolitic manuscripts on the GlagoLab portal.

#### **REFERENCES LIST**

- Australian Prints and Printmaking. Available at: https://www.printsandprintmaking.gov.au/ (accessed 11 December 2022).
- European Association for Digital Humanities (eadh). Available at: https://eadh.org/ (accessed 11 December 2022).
- GlagoLab Centar za istraživanje glagoljaštva Sveučilišta u Zadru. Portal and Digital Laboratory for Collaborative Research and Promotion of the Croatian Glagolitic Alphabet. Available at: https://glagolab.eindigo.net/ (accessed 11 December 2022).
- Glinka, Katrin, Christopher Pietsch, and Marian Dörk. 2017. "Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections". DHQ: Digital Humanities Quarterly 11 (2). Available at: http://digitalhumanities.org/dhq/vol/11/2/000290/000290.html (accessed 11 December 2022).
- Hall, Mark and David Walsh. 2021. "Exploring Digital Cultural Heritage through Browsing". In Information and Knowledge Organisation in Digital Humanities: Global Perspectives. Koraljka Golub and Ying-Hsang Liu, eds. London: Routledge, 261–284. Available at: https://doi.org/10.4324/9781003131816 (accessed 11 December 2022).
- Hibberd, Georgina. 2014. "Metaphors for Discovery: How Interfaces Shape Our Relationship with Library Collections". Available at: https://www. researchgate.net/publication/281230236\_Metaphors\_for\_discovery\_ how\_interfaces\_shape\_our\_relationship\_with\_library\_collections (accessed 11 December 2022).
- HistoGraph. Available at: http://histograph.eu/ (accessed 11 December 2022).
- History 503: Ancient Rome. Available at: https://tinyurl.com/hist503-timeline-sp18 (accessed 11 December 2022).
- Icelandic Saga Map. Available at: http://sagamap.hi.is/is/ (accessed 11 December 2022).
- Liu Chang, Ying-Hsang Liu, Jingjing Liu, and Ralf Bierig. 2021. "Search Interface Design and Evaluation". Foundations and Trends in Information Retrieval 15 (3-4): 243–416. Available at: http://dx.doi.org/10.1561/1500000073 (accessed 11 December 2022).
- NYPL Public Domain Release 2016 Visualisation. Available at: https://publicdomain.nypl.org/pd-visualization/ (accessed 11 December 2022).
- Rimmer, Jon, Claire Warwick, Ann Blandford, Jeremy Gow, and George Buchanan. 2008. "An Examination of the Physical and the Digital Qualities of Humanities Research". Information Processing & Management 44 (3): 1374–1392.
- Russell-Rose, Tony and Tyler Tate. 2013. Designing the Search Experience: The Information Architecture of Discovery. USA: Morgan Kaufmann.
- Victorian Illustrated Shakespeare Archive. Available at: http://www.shake-speareillustration.org/ (accessed 11 December 2022).

- White, Ryen and Resa A. Roth. 2009. Exploratory Search: Beyond the Query-Response Paradigm. [San Rafael, Calif.]: Morgan & Claypool Publishers.
- Wilson, Max L., Bill Kules, M. C. Schraefel, and Ben Shneiderman. 2010. "From Keyword Search to Exploration: Designing Future Search Interfaces for the Web". Foundations and Trends® in Web Science 2 (1): 1–97. Available at: http://dx.doi.org/10.1561/1800000003 (accessed 11 December 2022).
- Windhager, Florian, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. 2019. "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges". IEEE Transactions on Visualization and Computer Graphics 25 (6): 2311–2330. Available at: https://doi.org/10.1109/TVCG.2018.2830759 (accessed 11 December 2022).

# PARTICIPATORY MAPPING OF THE CULTURAL HERITAGE IN THE MOUNTAIN REGION OF GORSKI KOTAR IN CROATIA

# MARTINA PETRINOVIĆ

Croatian Society of Art Historians Zagreb, Croatia dpuh@inet.hr https://orcid.org/0000-0003-1972-0158

Art history, as a discipline within the humanities, has the potential to reach, impact, and engage the wider public with its research, and even involve the community in its research endeavours. The Croatian Society of Art Historians, a professional association whose goals include introducing the values of heritage to the local population, has undertaken a project involving art historians residing in a predominantly rural area of Croatia. The initiative of participatory mapping of the cultural heritage of Gorski Kotar engages the local community in a two-way exchange of knowledge. Participatory mapping is often used as a tool to help communities document and preserve their cultural heritage. By creating a free map together, community members can share their knowledge and perspectives, and work together to document their collective heritage. This can be an important step in the preservation and protection of cultural heritage for future generations.

Keywords: art history, topography, cultural heritage, conservation, open access

One of the key trends shaping the future of the humanities is the increasing importance of digital technology and the growing role of the digital humanities. This includes the use of digital tools and methods to study, preserve, and share cultural heritage, as well as the exploration of new forms of digital expression and creativity. The Croatian Society of Art Historians has undertaken an initiative for the digital documenting of cultural heritage in a remote rural area in Croatia with collected data in open access.

# PROJECT'S PURPOSE, OBJECTIVES, AND STAKEHOLDERS

Gorski Kotar (Figure 1) is the mountainous region in Croatia and a climatic barrier between the littoral and continental parts of the country. The population density of Gorski Kotar is low. With an area of about 1,300 square kilometres, it has one of the most negative natural increase rates recorded in the Republic of Croatia (Statistics 2021). Gorski Kotar has been regarded as a transit corridor between continental Croatia and north Adriatic ports (Jarec 2021).



Figure 1: Gorski Kotar area in Croatia (DPUH, 2022).

As such, the area of Gorski Kotar represents an insufficiently researched and properly valorised cultural space, as evidenced by the lack of conducted historical, cultural, or art history research and, consequently, literature (Crnić Novosel 2020). The attitude of the scholarly community has changed in recent years. Since the establishment of the Department of Art History at the Faculty of Humanities and Social Sciences, University of Rijeka, with their mentors' help, students from the area have been processing the neglected heritage in their works (Troha 2017, Tušek 2021). Their research shows that the region provides great potential for a more detailed understanding of its cultural heritage and a chance to preserve it from its decay. The cultural heritage of Gorski Kotar represents a true blank "white spot" on the map of Croatia and Europe. However, microeconomics postulates regard the white spot as "a place of unfilled potential – a blank slate or canvas that is waiting to be coloured in" (Kelwig 2022).

The Croatian Society of Art Historians is a non-profit organisation of art historians based in Zagreb, Croatia. Since its foundation in 1956, its members – professionals in the field of heritage protection, education, museology, research, conservation, and others – have been promoting the values of cultural heritage and the importance of visual arts. Art history is a discipline in the humanities that can have a vital role in providing the foundation for a society of knowledge and tolerance. Art history is free of religion-gender-nation bias. Research on artistic and cultural tangible heritage can encourage citizens to engage in critical thinking and participate in active dialogues. To bridge the gap between research and the wider public, the Croatian Society of Art Historians has set up and implemented many projects in the field of open science and open access.

Publishers' responsibility to society is to provide straightforward, meaningful, and direct communication of scientific research with the interested public. The basic tool for this process is the principles of open science used in correlation with long-term visionary actions advocating scientific excellence organised by professionals in the field (Petrinović 2020).

Humanities scholars and organisations can use several strategies to reach a wider public: developing and promoting outreach programmes that engage the public; utilising digital media platforms, such as social media, websites, and blogs, to share research findings, insights, and perspectives with the public; making research findings and information accessible to the general public through clear and concise writing and conducting research that is directly relevant to the needs and interests of specific communities, and engaging those communities in the research process.

In its seven decades of active scholarly publishing, and through collaboration with various institutions, organisations, and media outlets, the non-profit organisation of Croatian Art Historians has served as a prominent voice in cultural advocacy in Croatia, promoting the value of visual art and heritage. The scholarly community's reluctance to fully embrace openness is understandable, with recent warnings suggesting that the concept of open science may be naive and even lead to inequality. "Openings are also always closures and recognizing this is to recognize that the capacity for openness in open scholarship is unevenly distributed and its value and impact is unevenly weighted across the spectrum of participants and stakeholders" (Verhoeven 2022). However, we still did not witness the large-scale application of the notion of open science. That open access is not enough is obvious, and even commercial publishers, whose services are not free, are promoting their ideas on how to "spread the benefits of research and make the best use of research" (Gardner and Brassington 2022).

#### DOCUMENTATION OF HERITAGE

Detecting and documenting tangible cultural property is the first phase of undertaking the process of its valorisation. The process of systematic documentation of cultural and artistic heritage in Croatia was initiated by authorities in the early 20th century. The first researchers did not have an easy-going job and the people who were undertaking those duties as a rule involved engaged, enthusiastic, and committed investigators. Art historians Artur Schneider (1879-1946), Giuro Szabo (1875-1943), later Andela Horvat (1911-1985), and many others carried out extensive research of sites throughout Croatia. Their approach on the field was straightforward and resulted in an extensive listing of mainly architectural heritage and significant tangible artefacts accompanied by photographs, descriptions, and valorisation limited by the available information at the time. Their research was fundamental for heritage preservation in Croatia (Dulibić 2016, Špikić 2018, Armano 2011). The process of detecting, recording, data collecting, and creating topography and final valorisation is the same even nowadays. We can agree that there is no input without fieldwork and actual contact with the objects. The main interest of art history is objects and their reference to a human; therefore, this discipline provides opportunities for direct engagement with objects, images, and the built environment. But today the data researchers collect can be more manageable and userfriendly. Potential users of the data accompanied with metadata are researchers and the wider public.

The importance of this kind of fundamental research is underlined by one of the first UNESCO documents stating: "To ensure that the cultural and natural heritage is effectively recognised at all levels of planning, Member States should prepare maps and the fullest possible documentation covering the cultural and natural property in question" (UNESCO 1972). Furthermore, heritage is under constant threat from disasters, causing significant damage, particularly to architectural heritage, and posing risks to heritage structures. Unfortunately, we are witnessing the destructive forces of wars and natural hazards. After all catastrophes, a lot of energy is invested in reconstruction, and good sound documentation is vital for such processes. It is more critical for developing countries where often the case is poorly understudied.

Croatia today has an institutionalised and developed system of heritage protection on the governmental level. The Ministry of Culture and Media has developed a central register of cultural goods. The Geoportal of Cultural Property is a visual presentation of the protected cultural properties. Locations of great cultural importance are pinned on the cadastre.

The Geoportal of Cultural Property presents the scope of protection of the immovable cultural property under the competence of the Ministry of Culture and Media and therefore, any action is undertaken in compliance with the Act on the Protection and Preservation of Cultural Property (Geoportal 2022). The map of cultural property is also a product of research contributions from art historians such as Schneider, Szabo, and Horvat, who have dedicated themselves to fieldwork, as mentioned earlier.

#### PARTICIPATORY MAPPING: THE PRIMARY METHODOLOGY

The past twenty years and more have witnessed an explosion of participatory mapping initiatives throughout the world, in both developing and developed countries. "Changes have occurred since GIT&S have increasingly become accessible to civil society and graphic representations of space have been used as channels for two-way communication purposes to support social learning, dialogue, and negotiation processes" (Rambaldi 2005). Participatory mapping is simply the creation of maps by local communities. Participatory maps provide a valuable visual representation of what a community perceives as its place and the significant features within it. In our case, we are directing the conversation in the areas that we are interested in. Participatory mapping is a method of mapping in which members of a community are actively involved in the process of creating and maintaining a map. This approach is often used in situations where traditional mapping methods may not be effective, such as in remote or rural areas, or communities with little or no access to mapping technology. Participatory mapping can be used to document a variety of different things, including cultural and historical heritage. This can be especially important for communities that may be at risk of losing their cultural heritage due to various factors, such as urbanisation or changing social and economic conditions (La Frenierre 2008).

By involving community members in the mapping process, participatory mapping can also help to foster a sense of ownership and empowerment among community members. This can be an important factor in ensuring that the map is accurate and reflects the cultural heritage of the community. It can also help build stronger ties within the community and promote a sense of pride in the community's heritage. Focus groups of local inhabitants in our research very often were proud when we asked them to share information that they own and were willing to share their knowledge with a researcher (Figure 2).



Figure 2: Participatory mapping helps build stronger ties within the community and promotes a sense of pride in the community's heritage (DPUH, 2022).

The aim of our initiative is to record the diverse range of voices and not necessarily create one interpretation or "one story". Although the strategy is not envisaged as a traditional form of an educational programme, in the direct two-way communication we encountered true community and true trustful relationship between researchers and target groups and a remarkable two-way exchange of knowledge.

Documenting cultural landscapes or landscapes around cultural heritage objects has often been neglected because the surveying and mapping expenditure was too high. With modern surveying and visualisation techniques, landscape documentation and visualisation have become affordable and rewarding. And this is what our intention is. Although it is interesting and revealing – also it is time-consuming. We have to process a lot of information, that should function as curated content. While Google can be a useful tool for researchers, it is important to keep in mind that the information found on Google may not always be accurate or reliable. Researchers should critically evaluate the information they find and consider using additional sources to verify the accuracy of their findings. The Gorski Kotar project provides only accurate information.

#### PINNING VISUAL DATA ON A DIGITAL MAP

The Croatian Society of Art Historians initiated a project documenting tangible immovable and movable cultural and artistic heritage in Gorski Kotar as the first step in starting a new project. During the project initiation in 2022, the team of art historians established the project's purpose and the value it would deliver. Members of the team are art historians by profession (Martina Petrinović, Danko Dujmović, Daniel Štimac, Dejan Troha, Katarina Tušek, and Eugen Burić), privately connected to the area and familiar with the sites, which is one of the advantages of this research project. There is a substantial time-saving in scanning the area if a researcher is familiar with the area, exact locations, and potential tellers in the community that could be contacted. Also, there is a quality trust relationship between the researcher and the information holder that could end up in more accurate data collection. MARTINA PETRINOVIĆ

The team worked out the categories of mapping objects that include tangible cultural heritage (archaeological sites, churches, residential architecture, monuments, World War II memorials, industrial architecture, historical transportation infrastructure, documents, postcards, photographs, books, and industrial and graphic design products). The entries' locations are coordinated with the Register of Cultural Property of the Ministry of Culture and Media and the Geoportal of Cultural Property (Geoportal 2022) and pinned on the online web service Google Maps. The commercial application service provides free detailed information about the area's geographic features, and precise coordinates of a pinned entry, and offers aerial and satellite views of locations. Google Maps can be accessed through a web browser or mobile applications on smartphones and tablets. It is one of the most widely used mapping and navigation tools, providing detailed maps and information for millions of locations around the world. Google Maps entries can be enhanced with additional information and details to provide a comprehensive view of a location. Unfortunately, non-commercial platforms (e.g. OpenStreetMap) do not provide this kind of service and visibility. Despite the fact that commercial proprietary platforms maintain ownership of data, they possess a robust and all-encompassing user interface, which contributes to increased visibility and impact. This aspect is vital in attracting a larger user base and gathering additional data from other sources.



Figure 3: Illustration of various collected images (DPUH, 2022).

We are collecting visual data (Figure 3) – and we are feeding our map with those. At this stage we are not interpreting, we are collecting visual data. After entering the location in the digital map, data on literature, existing archival sources, useful links, and photographs of the entry have been added. As we pin an image (metadata added to the image or file) to an exact location on the map, data is attached and not detached from the community, and is subject to further use (Figure 4). Images uploaded to Google Maps are searchable through Google Lens. Google Lens is an image recognition tool that uses artificial intelligence to analyse and identify objects within images, including those uploaded to Google Maps.



Figure 4: Examples of entries on the digital map (DPUH, 2022).

# COMMUNITY AWARENESS OF HERITAGE

We intend to stimulate connections with the local municipalities to make the community aware of the values they inherit. Participatory mapping is the starting point for additional, more thorough, and interdisciplinary research. For a more detailed and professional historical-artistic stylistic analysis of each property of cultural importance, as well as their final systematisation and cataloguing according to the settlements in which they are located. MARTINA PETRINOVIĆ

It may seem that this area was forgotten by researchers and today it is facing substantial development of tourist infrastructure aggressive to the indigenous cultural landscape. In early 2023, the Institute for Physical Planning of Primorje-Gorski Kotar County Public Institution held an event aimed at increasing the local population's awareness of the unique identity of their settlements. The residents are well aware of the importance of safeguarding the visual identity of Gorski Kotar's settlements, as emphasised by the Tourist Board. The Institute is set to publish a comprehensive study that will serve as the foundation for developing spatial plans for all nine local administrative units within Gorski Kotar (Institute for Physical Planning of Primorje-Gorski Kotar County Public Institution 2023).

Cultural landscape refers to the physical and cultural environment created by human activities and interactions with the natural environment. It encompasses the built environment, such as buildings, infrastructure, and landscapes, as well as intangible elements, such as cultural practices, beliefs, and traditions. Cultural landscapes reflect the unique history, traditions, and values of the people who have inhabited them and are shaped by their interactions with the natural environment and each other. Cultural landscapes are dynamic and evolving, reflecting the ongoing cultural, social, and economic changes that take place over time. They can range from rural farmland to urban areas and can include parks, gardens, historic sites, and other places that hold cultural and historical significance (Dumbović Bilušić 2015). The new type of building clashes with the cultural landscape and indigenous traditional building.

# NAVIGATING PROJECT CHALLENGES: EXPLORING PITSTOPS ALONG THE WAY

The approach that we have chosen is a bottom-up approach focused on individual and micro, on the fundamental and physical, with the potential to identify opportunities through idiosyncrasies. This organic strategy with small beginnings that will eventually grow in complexity may result in a mess of systems without meeting the global purpose. But with planned optimisation of the process of collecting we try to simplify the project by dividing it into small and feasible tasks that can be executed easily. Every step of the process is followed by releasing the data on the internet. Due to being in a pilot version, the map we are currently inputting data into remains unavailable to the public. The decision regarding its availability will be determined by the team at a later time. In case the project is terminated or fails for any reason, the data will be available for future use, and the visibility of the results of our research will make the project sustainable if the intention and methodology of the initiative are picked up by other researchers. In this way, the possible pitfalls of a civil sector initiative could be bypassed. "New and fluid modes of organizing volunteering are emerging, characterised by blurry boundaries of membership and a lack of formally delegated authority. The potential for fluidly organised groups to act as 'mission belts' has been questioned due to their presumed ephemeral nature and incompatibility with the procedures of formal authorities" (Dyhrberg Højgaard 2022).

# DIGITAL INFRASTRUCTURE FOR ENGAGING DIVERSE AUDIENCES

Using a comparative method of visual data in fundamental research is the basic method used in the discipline of art history. Creating artistic topography in a digital map will speed up the process of research and process of communication with fellow researchers abroad, mainly in neighbouring countries Slovenia, Bosnia and Herzegovina, Austria, and Italy for researching the historical impacts on the area. Besides art historians, the project includes local inhabitants who hold accurate knowledge of their customary (and otherwise usually unrecorded) heritage objects and knowledge of their local environments, which can be expressed in easily understandable maps. This participatory method could stimulate connections with the local municipalities to make the community aware of the values they inherit.

The Croatian Society of Art Historians is an official partner of the European Commission initiative The New European Bauhaus. This initiative is the flagship of the EU efforts in the field of social inclusion, sustainability, and aesthetics. It will help to preserve European unique artefacts, cultural landscapes, historical sites, and buildings in the digital age. In the year 2022, the EU Commission launched a dialogue with Member States on the joint creation of a Collaborative Cloud for Cultural Heritage to help safeguard European cultural treasures through digital infrastructure. The Cloud aims to foster cooperation and co-creation among cultural, creative, and technology sectors. Hopefully, we will see results in that field. With an envisaged budget of €110 million until 2025 from Horizon Europe, the Collaborative Cloud will be a unique infrastructure that will

enable unprecedented transdisciplinary and large-scale collaboration between specialists, such as cultural heritage scholars, curators, archivists and conservators. It will provide cutting-edge technologies for digitizing artefacts, researching artworks, and documenting data, all of which will significantly advance and add a new digital dimension to cultural heritage preservation, conservation, and restoration. It aims to facilitate access to advanced technologies and remove barriers for smaller and remote institutions (Gabriel 2022).

The principles of the humanities are now embedded in the official European Commission policy and are the things that we can applaud, however, taking into account the inequality, we have to be aware that there are still infrastructure issues that have to be solved around Europe. Many populated areas in remote areas still have no high-speed internet connection or broadband. The area of Gorski Kotar is still facing severe depopulation, and we believe that a project like ours could contribute to pinning this area as an interesting and inviting living area – as through history it has been. We hope our initiative will evolve from a largely regional and Croatia-based project grounded in this remote area in the mountains - into a global information pool. Because of the substantial emigration from the countryside or the "rural exodus", the expatriate communities (e.g. in South America, Australia, and even nearby European countries) with the information our project will provide can explore their roots and understand their heritage. Today technology, visibility, and accessibility can enable reconnecting or reuniting lost ties.

Humanities scholars and especially learned societies can help to increase the visibility and impact of their work and reach new and diverse audiences with the insights and perspectives generated by humanities research.

# **REFERENCES LIST**

- Armano, Ermin. 2011. "Sjećanja na Anđelu Horvat". Peristil 54 (1): 25–32. Available at: https://hrcak.srce.hr/188738 (accessed 5 October 2022).
- Dumbović Bilušić, Dubravka. 2015. *Krajolik kao kulturno naslijeđe*. Zagreb: Ministarstvo kulture Republike Hrvatske and Uprava za zaštitu kulturne baštine.
- Crnić Novosel, Mirjana. 2020. "Gorski kotar u svjetlu raznovrsnih istraživanja". Problemi sjevernog Jadrana 18: 119–132. Available at: https://dx.doi. org/10.21857/9xn31co3ny (accessed 5 October 2022).
- Croatian Bureau of Statistics. 2021. "Natural Change in Population in the Republic of Croatia". Available at: https://podaci.dzs.hr/2022/en/29027 (accessed 5 October 2022).
- Dyhrberg Højgaard, Cristine. 2022. "New and Fluid Forms of Organizing Volunteering". In A Research Agenda for Civil Society. Kees Biekart and Alan Fowler, eds. Cheltenham: Edward Elgar Publishing, 189–200. Available at: https://doi.org/10.4337/9781800378155.00022 (accessed 5 October 2022).
- Dulibić, Ljerka, ed. 2016. Artur Schneider: 1879.–1946.: zbornik radova znanstveno-stručnog skupa Hrvatski povjesničari umjetnosti. Zagreb: Društvo povjesničara umjetnosti Hrvatske.
- Gabriel, Mariya. 2022. "Commission Joins Forces with Member States to Launch a Collaborative Cloud for Europe's Cultural Heritage". Press release. 21 June 2022. Brussels: European Commission.
- Gardner, Victoria and Laura Brassington. 2022. "Why Open Access Is Not Enough: Spreading the Benefits of Research". HEPI Policy Note 42. Oxford: Taylor & Francis.
- Institute for Physical Planning of Primorje-Gorski Kotar County Public Institution. 2023. The Spatial Identities of the Rural Settlements of Gorski Kotar Workshop Was Held in Delnice. Available at: https://zavod.pgz.hr/en/information/news?item=THE\_SPATIAL\_IDENTITIES\_OF\_THE\_RURAL\_ SETTLEMENTS\_OF\_GORSKI\_KOTOR\_WORKSHOP\_WAS\_HELD\_IN\_ DELNICE (accessed 5 October 2022).
- Jarec, Morana. 2021. Infrastrukturni projekti i društveno-kulturno oblikovanje prostora u Gorskome kotaru (Infrastructural Projects and Socio-Cultural Construction of Space in Gorski Kotar). Doctoral thesis. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu. Available at: https://doi.org/10.17234/ diss.2021.8597 (accessed 5 October 2022).
- Kelwig, Donny. 2022. "What Is White Space Analysis? The Ultimate Guide to Addressing Unmet Customer Needs". Available at: https://www.zendesk. com/blog/white-space-analysis/ (accessed 5 October 2022).
- La Frenierre, Jeff. 2008. "Mapping Heritage: A Participatory Technique for Identifying Tangible and Intangible Cultural Heritage". The International Journal of the Inclusive Museum 1 (1): 97–104.

- Ministry of Culture and Media. 2022. The Geoportal of Cultural Property. Available at: https://geoportal.kulturnadobra.hr/geoportal.html?lang=eng#/ (accessed 5 October 2022).
- Petrinović, Martina. 2020. "Scholar Reaching the Audience a Perspective of a Civil Society Sector Publisher in the Humanities". In ELPUB 2020 24th edition of the International Conference on Electronic Publishing. Available at: https://hal.science/hal-02544287 (accessed 5 October 2022).
- Rambaldi, Giacomo. 2005. "Who Owns the Map Legend?" URISA Journal 17 (1): 1–13. Available at: https://www.researchgate.net/publication/40117952\_ Who\_owns\_the\_Map\_Legend (accessed 5 October 2022).
- Špikić, Marko, ed. 2018. Gjuro Szabo: 1875.–1943.: zbornik radova znanstvenostručnog skupa Hrvatski povjesničari umjetnosti. Zagreb: Društvo povjesničara umjetnosti Hrvatske.
- Troha, Dejan. 2017. Sakralna arhitektura čabarskog kraja u 18. i 19. stoljeću. Master's thesis. Rijeka: Sveučilište u Rijeci, Filozofski fakultet u Rijeci. Available at: https://urn.nsk.hr/urn:nbn:hr:186:516744 (accessed 5 October 2022).
- Tušek, Katarina. 2021. Stambeno graditeljstvo brodmoravičkog kraja do prve polovine 19. stoljeća (Residential Architecture in the Brod-Moravice Region till the End of the First Half of 19th Century). Master's thesis. Rijeka: Sveučilište u Rijeci, Filozofski fakultet u Rijeci. Available at: https://urn. nsk.hr/urn:nbn:hr:186:580614 (accessed 5 October 2022).
- UNESCO. 1972. Convention Concerning the Protection of the World Cultural and Natural Heritage. 16 November 1972. The General Conference of the United Nations Educational, Scientific and Cultural Organization, meeting in Paris, at its seventeenth session, from 17 October to 21 November 1972.
- Verhoeven, Deb. 2022. "Scholarship in a Clopen World". Pop!Public. Open. Participatory 4. Available at: https://doi.org/10.54590/pop.2022.002 (accessed 5 October 2022).

# **SLOVENIAN HISTORICAL TOPOGRAPHY**: A MEDIEVALIST'S GAZETTEER

# MIHA SERUČNIK

ZRC SAZU Milko Kos Historical Institute Ljubljana, Slovenia miha.serucnik@zrc-sazu.si https://orcid.org/0000-0002-2784-0275

Place names are a part of a language that tends to change less than other parts of the vocabulary. Nevertheless, place names change in different ways and for different reasons over a sufficiently long period. The Middle Ages is a period far enough removed from the present to present many such problems to the students of that era. In regions where the (literate) elites used a different language than the (illiterate) peasant classes, the situation becomes even more complicated. On the territory of present-day Slovenia, three non-Slovenian written languages were in use until well into modern times: Latin, German, and Italian. A historical gazetteer is a tool that helps researchers of past eras to identify often unrecognisable place names and, in some cases, places that no longer exist. The Slovenian Historical Topography is a database of place names on the territory of present-day Slovenia. It is the result of cooperation between the teams at ZRC SAZU<sup>48</sup> Milko Kos Historical Institute and the Computer Systems Department of the Jožef Stefan Institute. It was designed from the beginning as a web application that allows researchers to search and view the data freely. Currently, it covers about half of the Slovenian territory. Recently, the web application has been revised and the database has been extended with new data. The changes include a refreshed user interface look as well as filtering options by type and time of toponym attestation. An important addition is also the introduction of line geographies for waterways and larger geographical features (mountain ranges, for example) which had been previously represented as points.

Keywords: the Middle Ages, gazetteer, topography, database, Slovenian lands, web map

<sup>&</sup>lt;sup>48</sup> Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti (Research Centre of the Slovenian Academy of Sciences and Arts).

#### INTRODUCTION

This paper is an extended version of a presentation given in October 2022 in Rijeka at the second DARIAH-HR Digital Humanities & Heritage conference.<sup>49</sup> It is a presentation of a long-term project to build a historical gazetteer for the Slovenian Middle Ages, conducted at the ZRC SAZU Milko Kos Historical Institute.<sup>50</sup>

Place names are a part of a language that tends to change less than other parts of the vocabulary. Nevertheless, place names change in different ways and for different reasons over a sufficiently long period of time. The Middle Ages is a period far enough removed from the present to present many such problems to the students of that era. In regions where the (literate) elites used a different language than the (illiterate) peasant classes, the situation becomes even more complicated. On the territory of present-day Slovenia, three non-Slovenian written languages were in use until well into modern times: Latin, German, and Italian. A historical gazetteer is a tool that helps researchers of past eras to identify often unrecognisable place names and, in some cases, places that no longer exist. The Slovenian Historical Topography is a database of place names on the territory of present-day Slovenia. It was designed from the beginning as a web application that allows researchers to freely search and view the data. Currently, it covers about half of the Slovenian territory. Recently the web application has been revised and the database has been extended with new data.

The paper offers a contextualisation of the problematics, describes the prehistory of the project, gives an overview of the current state of the project, and concludes with an outlook on future developments.

<sup>&</sup>lt;sup>49</sup> 2nd DARIAH-HR international conference Digital Humanities & Heritage, Rijeka, Croatia, 12–14 October 2022, Faculty of Humanities and Social Sciences in Rijeka, https://dhh.dariah.hr/home/.

<sup>&</sup>lt;sup>50</sup> "Slovenska historična topografija" ("Slovenian Historical Topography") or SHT for short. The enterprise has been made possible through two research projects financed by the Slovenian Research Agency: (1) "Slovenian Place Names in Time and Space (Historical Topography of Slovenia from the Middle Ages to the 19th Century)" (J6-4104), https://cris.cobiss.net/ecris/si/en/project/7016 and (2) "Toponomastical Heritage of Primorska Region" (L7-9424), https://cris.cobiss.net/ecris/si/en/project/17391. SHT team has been successful at acquiring a third project in 2022 named "Historical Topography of the Posavinje and Posotelje Regions" (L6-4606), https://www.zrc-sazu. si/en/programi-in-projekti/historicna-topografija-posavinja-in-posotelja. More on these projects is explained in the paper below.

# WHAT ARE GAZETTEERS AND WHY CREATE A HISTORICAL ONE?

Gazetteers are generally defined as indexes or dictionaries of place names. Entries are explained by location, geographical makeup, demography and so on. By extension, a historical gazetteer provides old names of settlements and geographical features that are no longer in use. Typically a historian's aid, such works must provide information on how the name of a locality changed over time and the sources from which the data are derived. Place names change over time. This is due to several factors: (1) changes in orthography, (2) changes in the language of the sources, and (3) administrative or political changes. The further back in time one reaches, the more varied the orthography one encounters. This is not surprising, since all living languages change over time. In addition, modern orthographic conventions did not emerge until the modern period as a result of the printed word. Before that, in the Middle Ages, the individualistic influences of the text producers - the scribes and their cultural background (language, schooling, etc.) and their (oral) informants - were all the more pronounced.

The language of the sources must also be taken into account. Slovenia as an independent nation-state has existed since 1991, before which it was part of Yugoslavia and its predecessor, the Kingdom of Serbs, Croats, and Slovenes (founded after the end of World War I). Even earlier, the Slovenian territory was part of the Habsburg Monarchy, which in turn was part of the Holy Roman Empire. The language of literacy in the Middle Ages at the southeastern border of the Holy Roman Empire was Latin and, from the end of the thirteenth century, German. On the northern Adriatic coast, the local variant of mediaeval Italian was in use (although Latin was used in the written sources of this region much longer than in the interior). Even later, in the early modern period and beyond, German and Italian persisted as almost exclusive languages of official records (with Latin remaining the language of the Church) well into the modern period. The Slovene language is found in administrative sources from the last two decades of the nineteenth century onward. On the other hand, the substrate from which the place names originated was Slavic or Slovene. The mediaeval scribes were usually not proficient in Slovene. One can point out some strategies that

scribes used when writing down Slovene place names. For example, they could ask for the meaning of the place name and translate it into German. Others preferred to try a transliteration of the phonetic image of the place name. Larger settlements, monasteries, and castles occur frequently in mediaeval texts, and their names have a more stable orthography. As a result, the historian is often confronted with place names that differ beyond recognition from those in use today.

In addition, one must consider the issue of localities that are no longer found on modern maps. Some localities become engulfed by neighbouring settlements and turn from a village into a part of a city, a name for a neighbourhood. In other cases, settlements were completely abandoned and disappeared from memory. Also in Slovenia, there are quite frequent cases of settlements that were renamed because their names had a religious meaning and this was not acceptable to the communist authorities after the end of World War II.

# A HISTORICAL GAZETTEER FOR THE SLOVENIAN MIDDLE AGES

The digital historical gazetteer under discussion is a project that owes a debt of gratitude to the work of three generations of historians from the twentieth century. The endeavour of compiling a gazetteer was first proposed in the early 1930s by Ljudmil Hauptmann, who at the time collaborated in the project of the Historical Atlas of the Austrian Alpine territories (see Hauptmann 1929). Hauptmann managed to get his colleague Milko Kos behind the idea and the two agreed to split the task. During the upheavals of World War II, Hauptmann lost all his materials connected to the gazetteer, and was greatly demoralised by this loss. Kos, on the other hand, was more fortunate and kept his materials. After the war, he became head of the newly founded Research Section for General and National History at the Slovenian Academy, which was later transformed into an independent institute. Kos kept working on the gazetteer project, preparing for its publication until his death in 1972. Afterwards, Pavle Blaznik, a colleague of Kos at the Historical Institute, took on the task of editing his legacy and published it in 1975 under the title Materials for the Historical Topography of Slovenia (Carniola until 1500). This was not a proper book, however, but was considered a working aid, a stepping stone on the path to a more complete edition that was to follow. The *Materials* were published in two volumes of 777 cyclostyled pages plus an index of 104 pages appearing in 250 copies. As it happened, this temporary solution remained the only gazetteer for the historical territory of Carniola (present-day central Slovenia) for the next forty years. Its rough form notwithstanding, the work proved to be an invaluable tool for historians and was seminal in several ways. Incidentally, the formulation from its title "historical topography" came to mean as much as "historical gazetteer" in the circles of Slovenian historians.

Meanwhile, Pavle Blaznik had been working since the mid-1960s on the topography of southern Styria (now the eastern part of Slovenia), a historical territory which was split between the Republic of Austria and the Kingdom of Serbs, Croats, and Slovenes (later Yugoslavia) after World War I. Similar to the work of Kos, the results of this work appeared in print posthumously in three parts between 1986 and 1989, but in proper book format. The third gazetteer for the Slovenian territory collated by Ivan Zelko covered the north-easternmost area of Prekmurje and appeared in print in 1982.

After the mid-1980s, a long hiatus in the production of historical gazetteers followed. This left the western part of Slovenia bordering Italy (approximately one third of the surface area) uncovered. Milko Kos had been collecting information about this territory in parallel with his work on Carniola. As this was originally the region assigned to his colleague Hauptmann, the index card catalogue pertaining to the western part of Slovenia was in a much more raw state than the materials concerning Carniola. Janez Šumrada, a member of the Historical Institute, continued to work on the catalogue and supplement the collected information, but the publication had to wait for better times.<sup>51</sup>

# **GOING DIGITAL**

In 2010, the late amateur historian and genealogist Leopold Mokotar approached the Milko Kos Historical Institute. Mr. Mokotar, who had

<sup>&</sup>lt;sup>51</sup> For the project pre-history see Kosi et al. 2021: pp. 7–9 and therein cited literature: Kos 1948; Kos 1954; Kos 1975.

at home scanned and OCR-ed (and corrected the OCR mistakes) Kos' "Historical Topography for Carniola", imagined the institute could, based on his digitisation, reprint the work, and handed it to our care. This coincided with the long-running sentiment at our institute that something should be done with the hitherto unpublished materials from Milko Kos' legacy catalogue. But, it was felt that something more than a simple revision and a reprint was called for. Those who had been using the "Topography", knew that it contained many deficiencies which called for careful critical revision. Besides, it was deemed it would be of much greater benefit if the gazetteer was accessible on the web, serving a wide audience and giving us the ability to constantly update the data. As a result, a team<sup>52</sup> was assembled and we submitted a project proposal the following year. The proposal was successful and we embarked on the first DH project led by the Historical Institute.<sup>53</sup>

The idea was to transfer the information from the digital text (then in Microsoft Word file format) to a relational database, georeference the place names and arrange a web page to access the data. We had some skills with Microsoft Access, so we modelled the database ourselves. Originally, we planned to enter data via forms but we have found this to be way too slow for the time frame of a three-year project. We were lucky to acquire help from a computer sciences researcher and part-time historian dr. Jurij Šilc, who was at the time employed at the Jožef Stefan Institute. He arranged (with the help of one of his students at the IJS) a text parser with the help of which we were able to mark up our text and transfer the data into our database.<sup>54</sup>

Quite a significant amount of work had to be carried out with classical historiographic methods. In the intervening decades since 1975, the historiographical field naturally progressed. Several critical editions of primary sources and historical studies have been published. This new

 $<sup>^{\</sup>rm 52}$  The original team consisted of colleagues from the Milko Kos Historical Institute, as well as colleagues from the Department of History at the Faculty of Arts in Ljubljana and Archives of Slovenia.

<sup>&</sup>lt;sup>53</sup> "Slovenian Place Names in Time and Space (Historical Topography of Slovenia from the Middle Ages to the 19th Century)" (Description is available at: https://zimk. zrc-sazu.si/en/programi-in-projekti/slovenian-place-names-in-time-and-spacehistorical-topography-of-slovenia-from).

<sup>&</sup>lt;sup>54</sup> The tools have been described in Šilc and Korošec 2020, an e-working report, deposited and accessible at the library of the Jožef Stefan Institute in Ljubljana.

information had to be consulted and the results were included in our project. As a result, many locations have been added to the gazetteer and many errors corrected. Moreover, one has to consider that primary sources change hands; due to the territorial pertinence principle archives exchange documents and this information needed to be included in our database.

The work of identifying/localising mediaeval toponyms has always heavily relied on historical maps in combination with Austro-Hungarian gazetteers from the nineteenth and early twentieth centuries. Milko Kos had access to the maps of the so-called "Third Military Survey", a series of maps covering the former Austro-Hungarian territory, drawn to the scale of 1:75,000.<sup>55</sup> Later generations of historians were able to also consult the maps of the "First Military Survey", a project dating from the second half of the eighteenth century and published as a source edition from 1995 onwards. These maps, alongside historical cadastral maps, are now readily available on the internet.<sup>56</sup>

Alongside the cartographic sources, the researchers on the project team consult the Theresianisches Kataster (the cadastre of Maria Theresia) and the *urbaria* (rent-rolls) of manorial estates from the Slovenian territory. The consultation of these sources enables the creation of timelines for the morphologic changes of placenames and bridging the temporal gap between modernity and carters from the Middle Ages. It is also possible to track real-estates changing hands etc. where the level of sources' preservation is sufficient. Through such cross-referencing, the researchers have been able to double-check the localisations of toponyms. At the end of the first project, the database included over 3,500 primary records. Over one-fifth (835) of erroneous localisations from the original gazetteer have been corrected (Kosi et al. 2021: 9–14).

The resulting web page was centred around a list view of toponyms and a web map. The technologies employed were Google Maps with some custom JavaScript, and PHP scripts for communicating with the database and a MySQL database. Functionally the users were able to

<sup>&</sup>lt;sup>55</sup> Spezialkarte 1:75.000, Das Militärgeographische Institut, Wien 1872–1918.

<sup>&</sup>lt;sup>56</sup> See project "Arcanum Maps", formerly named "Mapire": https://maps.arcanum. com/en/.

browse all records on the map or on the list and filter them by toponym. As mentioned above, the dataset was limited to the territory of central Slovenia. On clicking a specific marker on the map or a record in the list, a detailed view of the historical names of the location (we dubbed them paleonyms) opened to the right of the screen (Figure 1).



Figure 1: A typical toponym record. The left pane holds a list of toponyms. The centre pane is reserved for the map. The right-hand pane displays paleonyms of the selected toponym. Paleonyms contain further details.

In 2018, a follow-up project began with the aim of adding the region of Primorska to our digital gazetteer. This time the team included members from the Jožef Stefan Institute from the very beginning. As with the first project, we based our efforts around the materials from the legacy of Milko Kos, with the distinction that we were bringing to light hitherto unpublished data.

First, the index cards were scanned and OCR-ed. Then we proceeded with mark-up and parsing. Again, the data were subject to thorough double-checking by means of studying old and new critical editions. Despite the COVID-19 lockdowns, which surprised us in the middle of the project, the team also successfully conducted a couple of field trips to archives in Italy.

Conceptually and technically the project was a case of building on previous experiences as well as introducing changes and new functionalities. The text parser and the database schema have been overhauled. Most of the software components have been retained (MySQL as database, PHP for server programming), but we have moved from Google Maps to a combination of Open Layers and OpenStreetMap. On the front-end side, much effort went into conceptualising data filters that would make sense to the user. In its second incarnation, the gazetteer offers filtering by toponym, region, type categories, as well as by year.

The temporal filter is provided by a slider through which the user can specify the time interval in which they are interested. The closest resolution of the time filter is a single year. This includes reduced precision dates (e.g. only the year is known) and uncertain dates (approximate date, *terminus ante quem*, *terminus post quem*) which are all resolved to a single year number by the application. The filter acquires the years dynamically from the map features at the time of page loading. Map features are, on the other hand, created from the database records and saved as a JSON file on the server.



Figure 2: Current extent of the gazetteer.

The gazetteer also includes a rich descriptive typology, inherited from the work of Milko Kos and his successors. The types are displayed in the localities' details. Because this typology (currently) comprises 84 distinct terms, it has been reduced to five categories (settlements, waterways, mountains, broader geographic features, and built structures) for feature filtering.

The last filter is called "regions". These correspond to areas included in the first and second projects and are not to be equated with historical territories or present-day administrative units. Such an arrangement was decided upon due to practical reasons. The entire project is organised about the modern national territory of Slovenia. Its boundaries in most cases do not correspond to borders from the mediaeval period. Furthermore, the historical territories have formed gradually from the High mediaeval period onward and their development was in no way homogeneous. Thus, trying to provide historically accurate political boundaries would prove a complicated endeavour and, in the end, subject to "editorial" decisions. The two regions currently offered to the user are Carniola and the Primorska region (Littoral).

As indicated above, all filters are executed in memory on the client (JavaScript), thus avoiding repeated queries of the database and risking slowing down the application. Even so, when the majority of the features are being removed or added to the view, some lag is incurred, and we are looking into strategies for mitigating this.

Aside from the technical changes, the web page received a visual upgrade as well. While retaining the basic structure of the interface, we modernised the CSS code. The typography adjusts the size for higher resolution displays, a project logo was added and the added filters had to be appropriately placed. As a whole, we hope we have managed to provide a refreshed look and a frustration-free user experience.

Another significant change was introduced on the map. In the original version, all features were represented with point geometries (markers). With the recent version, polylines have been added to represent waterways and polygons for the lakes.

The gazetteer currently holds 5574 primary records (toponyms) from Carniola and the Primorska. It is worth also mentioning that the

study of sources for the Primorska resulted in the addition of 620 further toponyms in the Carniola region. A query of paleonyms currently yields 25 289 records.

# DISCUSSION AND OUTLOOKS

The Slovenian historical topography has started as a digitisation effort and this is strongly reflected in the way we have tackled the question of structuring our data. Secondly, it has been our ambition to provide a practical tool for researchers to browse and query the data over the web. Thirdly, the task of our gazetteer is to provide toponomastical information on the territory which would otherwise remain only sporadically and unsystematically covered in gazetteers not based in Slovenia. As such, our team has mainly pursued practical concerns on how to think about our project matter. Nonetheless, one can make a few more formal statements to describe our project.

The gazetteer is based on sources that are legal (charters) and administrative (urbaria) in nature. As such, it contains non-fictional places only. The majority of these are settlements (villages, towns, market towns), important buildings (castles, churches, monasteries), but also geographical features (rivers, mountain ridges, bogs, etc.). In general, we do not cover administrative or political entities, though in some cases the territories of a seigneurial morphed into names of (micro) regions. Thus our gazetteer could be said to include some instances of fiat objects. Still, we have decided at an early stage that modelling relations such as places belonging to a political entity (e.g. county) is beyond the scope of our project.<sup>57</sup>

On the question of temporal disambiguation, we have always been focused on our primary sources which are predominantly charters. In cases where the documents themselves do not contain information about the date of their creation, we rely on the findings of historiogra-

<sup>&</sup>lt;sup>57</sup> The distinction between physical and fiat objects is taken from Schneider et al. 2019. The paper provides an in-depth discussion of the problem of historical gazetteers and investigates several issues connected to the mediaeval period, such as the difficulties of defining political and administrative entities – the vagueness of their territorial extent, problems of change in time and so on.

phy. Thus every paleonym is equipped with at least one time stamp at which it is valid.  $^{\rm 58}$ 

So, what course of action lies ahead? In 2022, our, by now well-knit team, was successful in obtaining a third project, this time aimed at converting the historical gazetteer for Slovenian Styria by Pavle Blaznik into digital format. Blaznik's work represents the largest body of text our team grappled with. Due to different geographic characteristics, this area exhibits far more fragmented settlement patterns than central and western Slovenia. It also lacks large sparsely settled areas such as the Julian Alps. These factors contribute to a significantly greater number of toponyms. As a consequence, we have resolved to split Blaznik's gazetteer in half and concentrate on the western and southern parts of the territory covered in his book,<sup>59</sup> leaving the other half for a later project.

The final goal is, of course, to cover the entire territory of Slovenia, as well as the bilingual areas in neighbouring countries with Slovenian minorities.

Aside from extending the gazetteer's territorial coverage, we are planning to provide a translation of the interface which at the moment still offers only Slovenian language. Our other ambitions include an API for querying the gazetteer programmatically; most likely serving query results in JSON format.

Also, an advanced search mode interface for the site is long overdue. With this, we would like to provide the user with more full-featured access to the information in our database. As we provide source references for all paleonyms included, the gazetteer represents the most accurate and the most inclusive collection of primary sources and critical editions for the Slovenian Middle Ages to date. The addition of bespoke views will be without doubt of great help to any user, professional or amateur.

Finally, we need to address the question of the relationship of our gazetteer to projects aggregating sources such as "Pelagios" and

<sup>&</sup>lt;sup>58</sup> And some of the paleonyms with many timestamps. For a discussion of temporal disambiguation, see: Schneider et al. 2019: 7.

 $<sup>^{\</sup>rm 59}$  This territory largely corresponds to the area under the control of the Counts of Cilli from the fourteenth century onward.

"World Historical Gazetteer". As mentioned earlier, Slovenian territory is only poorly covered regarding historical placenames and offering our information to an international project would certainly be of value. Any such endeavour would of course entail meeting the requirements of the aggregator. While such an ambition is not one of our immediate priorities, we could describe it as a medium-term possibility.

# ACKNOWLEDGEMENTS

The work presented in this paper was made possible through three projects financed by the Slovenian Research Agency,<sup>60</sup> funds provided by the Slovenian Academy of Sciences and Arts, and supported by DARIAH-SI research infrastructure.

# **REFERENCES LIST**

- Arcanum maps. Available at: https://maps.arcanum.com/en/ (accessed 5 April 2023).
- Blaznik, Pavle and Darja Mihelič. 1986–1989. Historična topografija Slovenije II. Slovenska Štajerska in jugoslovanski del Koroške, do leta 1500 (3 deli). Maribor: Zal. Obzorja.
- Gemeindelexikon von Krain. Bearbeitet auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1900. 1905. Wien: Herausgegeben von der k. k. statistischen Zentralkommission.
- Hauptmann, L[j]udmil. 1939. "Krain". Erlauterungen zum Historischen Atlas der österreichischen Alpenländer I. 4: 309–484.
- Kos, Milko. 1948. Srednjeveški urbarji za Slovenijo 2. Urbarji Slovenskega primorja 1. Ljubljana: Akademija znanosti in umetnosti.
- Kos, Milko. 1954. Srednjeveški urbarji za Slovenijo 3. Urbarji Slovenskega primorja 2. Ljubljana: Slovenska akademija znanosti in umetnosti.
- Kos, Milko. 1975. Gradivo za historično topografijo Slovenije: (za Kranjsko do leta 1500). Ljubljana: Inštitut za občo in narodno zgodovino Slovenske akademije znanosti in umetnosti.
- Kosi, Miha, Matjaž Bizjak, Miha Seručnik, and Jurij Šilc. 2015–2024. "SHT Slovenska historična topografija". Available at: https://topografija.zrcsazu.si (accessed 5 April 2023).
- Kosi, Miha, Matjaž Bizjak, Miha Seručnik, and Jurij Šilc. 2021. Historična topografija Kranjske (do leta 1500). Druga dopolnjena izdaja. Ljubljana: Založba ZRC.

<sup>&</sup>lt;sup>60</sup> Recently renamed to Slovenian Research and Innovation Agency.

- Orts-Repertorium des Herzogthums Krain. Auf Grundlage der Volkszählung vom 31. Dezember 1869 / Imenik krajev vojvodine Kranjske. Sestavljen na podlagi ljudskega štetja od 32. decembra l. 1869. 1874. Laibach/Ljubljana: Druck und Verlag von Ign. Kleinmayr & Fed. Bamberg.
- Schneider, Philipp, Jim Jones, Torsten Hiltmann, and Tomi Kauppinen. 2018. "Challenge-Derived Design Principles for a Semantic Gazetteer for Medieval and Early Modern Places". Semantic Web Journal (1): 1–20. Available at: https://semantic-web-journal.net/content/challenge-derived-designprinciples-semantic-gazetteer-medieval-and-early-modern-places-1 (accessed 5 April 2023).
- Special-Orts-Repertorium des Herzogthums Krain/Obširen imenik krajev na Kranjskem.1884.Wien:AlfredHölderk.k.Hof-undUniversitäts-Buchhändler.
- Special-Orts-Repertorium von Krain. Neubearbeitung auf Grund der Ergebnisse der Volkszählung vom 31. Dezember 1890/Specijalni Repertorij krajev na Kranjskem. Na novo predelan po rezultatih popisa ljudstva dne 31. decembra 1890. 1894. Wien: Alfred Hölder k. u. k. Hof- und Universitäts-Buchhändler.

Spezialkarte 1:75.000. 1872–1918. Wien: Militärgeographische Institut.

- Šilc, Jurij and Peter Korošec. 2020. Programska orodja in pripomočki za pripravo podatkov in vnos ter izpis v oz. Iz podatkovne zbirke Historična topografija Slovenije. Report.
- Zelko, Ivan. 1982. Historična topografija Slovenije I. Prekmurje do leta 1500. Murska Sobota: Pomurska založba.

Founder of the Digital Humanities & Heritage conference: DARIAH-HR Conference website: https://dhh.dariah.hr/

**COMMITTEES:** 

2021

**ORGANISING COMMITTEE:** 

Koraljka Kuzman Šlogar Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

**Anamarija Žugić Borić** Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

Ana Antolković Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

Irena Miholić Institute of Ethnology and Folklore Research, Croatia

# Marijana Tomić

University of Zadar, Department of Information Sciences, Centre for Research in Glagolitism, Croatia

#### Zrinka Džoić

University of Zadar, Department of Information Sciences, Croatia

#### Laura Grzunov

University of Zadar, Department of Information Sciences, Croatia

# Nevenka Kalebić

University of Zadar, Department of Information Sciences, Croatia

#### Drahomira Cupar

University of Zadar, Department of Information Sciences, Croatia

**Vlatka Lemić** University of Zagreb / ICARUS Hrvatska, Croatia

**Saša Potočnjak** University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

PROGRAMME COMMITTEE:

Koraljka Kuzman Šlogar Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

# Marijana Tomić

University of Zadar, Department of Information Sciences, Centre for Research in Glagolitism, Croatia

# Saša Potočnjak

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Cécile Chantraine Braillon

La Rochelle University, France

#### Antoine Doucet

La Rochelle University, L3i: IT, Image and Interaction Laboratory, France

#### Anne Gilliland

Department of Information Studies, Graduate School of Education & Information Studies, UCLA, Los Angeles, USA

#### Vlatka Lemić

University of Zagreb / ICARUS Hrvatska, Croatia

#### **Benedikt Perak**

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

#### **Erich Renhart**

University of Graz, VESTIGIA Manuscript Research Center, Austria

#### Anna Maria Tammaro

University of Parma, Department of Cultural Heritage, Section Librarianship, Italy

#### Mirna Willer

retired from the University of Zadar, Department of Information Sciences, Croatia

# Tihomir Živić

Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek, Croatia

#### 2022

**ORGANISING COMMITTEE:** 

# Koraljka Kuzman Šlogar

Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

# Antonia Hladilo Duspara

Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

# Anamarija Žugić Borić

Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

# Irena Miholić

Institute of Ethnology and Folklore Research, Croatia

# **Benedikt Perak**

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Kosana Jovanović

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Barbara Španjol-Pandelo

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Luca Malatesti

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Vlatka Lemić

University of Zagreb / ICARUS Hrvatska, Croatia

# Marijana Tomić

University of Zadar, Department of Information Sciences, Centre for Research in Glagolitism, Croatia

### **PROGRAMME COMMITTEE:**

# Koraljka Kuzman Šlogar

Institute of Ethnology and Folklore Research, DARIAH-HR, Croatia

# Kosana Jovanović

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Antonija Primorac

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Barbara Španjol-Pandelo

University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

# Cécile Chantraine Braillon

La Rochelle University, France

# Josip Ćirić

University of Zadar, Department of Information Sciences, Croatia

#### Antoine Doucet

La Rochelle University, L3i: IT, Image and Interaction Laboratory, France

#### Anne Gilliland

UCLA, Graduate School of Education & Information Studies, Department of Information Studies, Los Angeles, USA

#### Koraljka Golub

Linnaeus University, Faculty of Arts and Humanities, Department of Cultural Sciences, Sweden

**Juan Gomis Coloma** Catholic University of Valencia, Spain

**Simona Grigaliūnienė** Klaipėda University, Klaipėda, Lithuania

**Ulrike Henny-Krahmer** University of Rostock, Germany

Meike Klettke University of Rostock, Institute for Informatics, Germany

**Vlatka Lemić** University of Zagreb / ICARUS Hrvatska, Croatia

**Erich Renhart** University of Graz, VESTIGIA Manuscript Research Centre, Austria

**Dalia Marija Stančienė** Klaipėda University, Lithuania

#### **Anna Maria Tammaro** University of Parma, Department of Cultural Heritage, Section Librarianship, Italy

**Marijana Tomić** University of Zadar, Department of Information Sciences, Centre for Research in Glagolitism, Croatia

Georg Vogeler

Austrian Centre for Digital Humanities, Graz, Austria

**Mirna Willer** retired from the University of Zadar, Department of Information Sciences, Croatia

Ulrike Wuttke University of Applied Sciences Potsdam, Germany

#### Tihomir Živić

Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek, Croatia
**REVIEWERS OF THE EDITION:** Boris Bosančić, Cécile Chantraine Braillon

**PAPER REVIEWERS:** Petra Bago, Melanija Belaj, Cécile Chantraine Braillon, Sanja Đurin, Anne Gilliland, Koraljka Golub, Maša Grdešić, Alba Irollo, Agata Juniku, Vlatka Lemić, Sanja Lončar, Petra Matović, Ana Mihaljević, Željka Miklošević, Benedikt Perak, Iva Pleše, Mojca Ramšak, Miha Seručnik, Dalia Marija Stančienė, Marijana Tomić, Željko Trbušić, Georg Vogeler, Jure Volčjak, Ana Vukadin, Mirna Willer, Ulrike Wuttke, Goran Zlodi, Tihomir Živić

The e-edition is freely available under the Creative Commons CC BY-NC-ND 4.0 International License.