



Collecting, Interpreting, and Digital Processing of Historical Data

29 September, DHI

Session 1

Maciej Janowski • *Do Facts Give us Knowledge?*

Krzysztof Ostafin • *From Census Tables to Spatial Data: Digitizing and Mapping Socio-Economic Statistics in Galicia (1857–1910)*

In the second half of the nineteenth century, statistical surveys and population censuses became key instruments for describing and measuring society in the Habsburg Empire. Galicia, a multiethnic crown land characterized by significant regional disparities, was the subject of numerous statistical investigations covering population, economic activity, education, religion, and social infrastructure. These statistical practices played an important role in shaping the quantitative knowledge used to measure and interpret society in the province.

This study presents a digital reconstruction and analysis of an extensive set of socio-economic data for Galicia covering the period from 1857 to 1910. The dataset integrates information derived from population censuses, statistical publications, and historical topographic and economic maps. The data were reconstructed at two spatial levels: municipalities and administrative districts. The collected variables include population size and structure, nationality and religion, occupational structure, livestock numbers, educational indicators, and the presence of economic institutions and social infrastructure.

The study focuses on transforming historical statistical tables into a structured digital database linked to historical administrative units. Using GIS-based methods, it demonstrates how the digital processing of historical statistics enables new interpretations of the spatial diversity of Galician society and economy at the turn of the twentieth century.

Anne Knowles • *Historical GIS as a Framework and Springboard for Spatial Stories of the Holocaust*

Over the past several decades, probing scholarship has discussed the limitations of GIS and conventional cartography for humanistic research and for communicating geographical understanding

to a broader public. At the same time, cartographers within and outside of the academy have developed representational approaches that seek to express lived experience in the world. This paper takes, as its starting point, a gap between those bodies of work: namely, how GIS can be both a framework for analysis and a means of bringing empirical rigor to visual storytelling that loosens the bonds of the coordinate grid. My efforts to fill that gap have employed substantial, manually created historical GIS databases of SS camps and Jewish ghettos in German-occupied Eastern Europe along with experimental collaborations with creative, analytical cartographers and graphic artists. To explain the process, I will outline the potential and limitations my research team encountered while designing the ghettos database and during data entry. A series of GIS-generated maps will demonstrate the data's analytical potential for revealing previously unspecified spatio-temporal dynamics of community destruction during the fraught history of ghettoization, as well as regional comparisons that highlight some of the complex variations in that history. Then, in a series of maps and hand-drawn graphics that incorporate uncertainty, ambiguity, and emotions of Jews' accounts of their experiences, I will suggest how the limits of GIS became a springboard for less constrained, more expressive, but still empirically grounded visualizations of the spatial history of the Holocaust.

Dariusz Gierczak • *xTree Geografika – a Data Model for Spatial Search for Geographical Names and Administrative Units*

The current borders of administrative divisions in most countries are relatively well documented. However, as soon as data from the past is needed, the available information becomes significantly scarcer, especially when it comes to tracking changes in borders on maps. Older data often disappears from online directories as soon as an update is made. Data from the pre-digitalization era is, if available at all, usually limited to higher-level administrative units. When historical data on a specific settlement or administrative unit is needed, the available data typically only extends up to the most recent administrative reform. Searching for older data is laborious, especially since it is scattered across various sources. Every war or annexation brought changes to the administrative system. Furthermore, place or administrative unit names also changed for other reasons.

Based on its own expertise and lively exchanges with partner institutions the Herder Institute has developed a structure for data organization and searching by spatial criteria for individual places and administrative units, including their historical development. Access to the data will be facilitated through the controlled vocabulary xTree.geografika, which is the focus of this presentation.

Morgane Labbé • *Measuring the Polish Nation before the Polish State: Reading Polish Statistics as a database compiled from the German, Russian and Austrian Statistical Bureaucracies*

Digitalized databases open up unexpected promises of access to an abundance of statistics at the local or even the household level, for historians who have long struggled with scattered sources, estimated data, and incomplete statistical series. But they also bring new methodological and even epistemological challenges. They reinforce the fiction of data realism and anachronism, and thus weaken our critical approach to quantitative sources. In this respect, looking to the statistical practices of the past can be instructive and the case of Poland highly relevant.

The rise of statistics, as both a science and an administrative practice, is linked to the emergence of the territorial and centralized state, backed by an extensive bureaucracy. In line with this approach, there has been a proliferation of studies examining the role of statistics in the formation of the nation-state and in the consolidation of imperial and colonial conquests. Between the late 18th century and 1918, there was no Polish state, and consequently no Polish statistics; however, the three Partitioning Powers conducted increasingly comprehensive and detailed censuses of the populations, including ethnic-national and religious minorities.

While they had to guide discriminatory and assimilationist policies, Polish activists, rather to contest and oppose them, appropriated these official sources to build their own Polish statistics, an alternative one against the state.

In my presentation I will argue that the Polish activists used the census statistics as databases, breaking down them in units and data modules into new combinations and rearrangements then new statistical tables backing their national claims, likewise does nowadays any user of digitalized database, making up his/her own file from dematerialized survey.

Philipp Kröger • *Measuring the Unmeasurable: Statistics, National Indifference, and Population Policies in the German-Polish Borderlands around 1900*

The question of how ethnicity could be recorded statistically occupied Prussian and German statisticians from the mid-nineteenth century onward. According to these scholars, the precise boundaries of a German nation-state could be delineated through statistical data. However, when ethnicity was first recorded in censuses via mother tongue following the founding of the German Empire, the results – particularly along the eastern border – revealed the opposite: ambivalence and national indifference rather than clear-cut order. In other words, it became apparent that ethnicity could not be measured without contradiction. This did not lead statisticians to abandon ethnicity as a statistical category; rather, it gave rise to a debate that lasted for decades over how it should be recorded.

Building on these observations, my paper advances two arguments. First, the Prussian-German example invites a critical perspective on nationality statistics: ethnicity is less a quantifiable entity in the world than a perspective on it (Rogers Brubaker). Second, even if ethnicity is assumed to be measurable without contradiction, the data produced cannot be separated from their political implications. As early as 1900, nationality statistics served not merely to describe the world but to shape it in accordance with ethno-nationalist conceptions.

Humphrey Southall • *Representing a changing nation: working geographically with the British censuses, 1801 to 2021*

The United Kingdom has been systematically documenting its changing society, economy and localities through censuses every ten years since 1801, missing only 1941. These have generated an enormous body of data in three main forms. Firstly, the original information about individual people, only available once the census carried out house-by-house surveys, and kept secret for a hundred years, so currently available only for 1851 to 1921. Secondly, very detailed "born-digital" statistics available from 1971 onwards. Thirdly, printed statistical reports, almost the only data available for 1801 to 1841, and 1931 to 1961. Using this data to analyse long-run trends poses several challenges. Firstly, systematic analysis generally requires very large scale digitisation. Secondly, there have been very large changes in the geographical frameworks used by the census, and these were well-documented only from the 1970s onwards: mapping earlier census geographies has been the central task of the Great Britain Historical GIS. Thirdly, both the questions asked by the census and how the answers were classified have necessarily changed greatly. This last issue will be explored through data on occupations and industry, enabling mapping of the changing economy.

Piotr Guzowski, Radosław Poniat, Marzena Liedke • *The Fifth Russian Revision of 1795 as a Source for Demographic Research into the Population of the Eastern Territories of the Polish-Lithuanian Commonwealth*

The Fifth Russian Revision of 1795 was a typical tool of the Russian bureaucracy for assessing the tax and military potential of the country's population (the first was conducted in 1719, the last in 1857). In the context of the population of the Polish-Lithuanian Commonwealth, it is of particular significance, as for most areas the censuses conducted in 1789–92—that is, prior to the Second and Third Partitions of the country—have not survived, whereas the Russian source has survived to the present day in Russian, Lithuanian, Belarusian and Ukrainian archives.

It provides a good insight into the state and structure of the population, including its social and occupational structure (mainly in rural areas), and allows for the analysis of the family and its household. The authors of the paper will highlight the research opportunities associated with the source, its strengths and weaknesses, and its regional specificity.

Keynote

Bartosz Ogórek • *A Proxy or a Delusion? Variable Design in Research on Historical Human Capital*

Like other social scientists, historians frequently have to work under conditions not of their own choosing. Source scarcity and the lack of direct access to past actors force them to bridge the gap between fragmentary evidence and the phenomena they wish to study. This often requires linking theoretical concepts to measures and classifications that can be constructed from the available sources (Adcock and Collier 2001). In other words, “turn[ing] proper names into variables and words into numbers” (Kreuzer 2010). Yet such proxies vary greatly in their validity. Focusing on research on historical human capital, this paper reviews several strategies of proxy construction and critically assesses what these measures actually capture by confronting alternative approximations with one another. Drawing on civil marriage certificates from the southern Polish lands in the first half of the nineteenth century, it discusses three commonly used indicators: age heaping as a measure of numeracy, but also of administrative proficiency; the ability to sign one’s own name as a proxy for literacy; and signature quality as an indicator of writing fluency (Tollnek and Baten 2024; Soubeyroux 1998; Cressy 2020).

30 September, DHI

Keynote

Catherine Gibson, Anton Kotenko • *Enumerating a Multispecies Romanov Empire: The 1897 Tartu Dog Tax Register and Canine Demography*

The year 1897 in the history of the Romanov Empire is conventionally associated with the first all-imperial (human) population census. This article examines another act of enumeration conducted that year: Tartu’s municipal registration of taxable dogs. Situating the register within the wider history of dog taxation across the empire, it traces how city authorities transformed dogs into objects of fiscal, sanitary, and spatial regulation. By compiling the register, officials connected animals to legally responsible owners and addresses, distinguished taxable dogs from exempt or ownerless ones, and

made compliance visible through annually renewed tags. Classifying dogs as owned or ownerless, working or companion, had material consequences for how they could live and move through the city. The resulting register, recording details of 1,436 dogs, also constitutes a unique dataset for reconstructing Tartu's canine demography, the animals' characteristics, names, and spatial distribution, and relating it to the social urban geography documented by the human census. Whereas scholarship on imperial censuses has shown how enumeration simultaneously documented and administratively constituted the Romanov Empire's social and ethnic heterogeneity, this presentation extends this argument beyond human populations, demonstrating how administrative records can reveal – and helped produce – the multispecies diversity of imperial urban life.

Session 3

Vojtěch Pojar • *Measuring the Nation's Body in the Age of Extremes: Mass Gymnastics, Physical Anthropology, and Nationalism, 1910s–1990s*

Nationalist mass gymnastics was a ubiquitous phenomenon in twentieth-century East Central Europe, closely linked to organic conceptions of nationhood. Designed to shape and display a collective, able “national body” in motion, its discourses and practices were rapidly medicalized, increasingly permeated by eugenic thought, and progressively reliant on physical anthropologists for anthropometric data. In turn, these settings provided anthropologists with opportunities to conduct regular measurements on large populations of research subjects. The Bohemian Lands, and later Czechoslovakia, constitute a paradigmatic case of this alliance between mass gymnastics, nationalism, and science.

This study examines large-scale anthropological measurements periodically conducted within the Sokol gymnastics movement and, later, at state-socialist Spartakiads. It argues that, despite major political ruptures, key discourses and practices persisted across three contexts: late imperial Austria-Hungary, the interwar nationalizing state, and the state-socialist dictatorship in Czechoslovakia (1948–1989). The paper shows how racialized concerns with the health, efficiency, and distinctiveness of the national body, along with biopolitical ambitions to optimize it, remained remarkably stable while adapting to shifting ideological frameworks.

Mikołaj Szołtysek • *Measuring Difference: Family Systems in East-Central Europe between Ideology and Data*

For much of the late nineteenth and twentieth centuries, scholarship portrayed East-Central European family systems as fundamentally distinct from those of Western Europe—described as “eastern,” “alien,” or “non-European”—a narrative sustained more by ideological presupposition than by systematic measurement. I address this point from two complementary perspectives. First, I show how these claims can be tested using an unprecedented corpus of census-like archival sources from the late eighteenth century (immediately prior to the dissolution of the Polish-Lithuanian Commonwealth) transformed into a harmonized, interoperable database. Drawing on my book *Rethinking East-Central Europe: Family Systems and Co-Residence in the Polish-Lithuanian Commonwealth* (Peter Lang, 2015), I demonstrate the data’s capacity to generate rigorous demographic measures that precisely quantify and spatially map differences and similarities. Second, by combining digital humanities workflows (digitisation, harmonization, and microdata construction) with critical historiography, I show that data-driven reconstruction not only overturns speculative typologies but also enables a critical reappraisal of East-Central Europe as a historical region—problematising monolithic labels and grounding debates about regional connectedness and the developmental divergences that the soon-to-decline Polish-Lithuanian Commonwealth would have confronted had it not entered decline.

Tomasz M. Jankowski • *Early Census Taking of the Jews in East-Central Europe*

The census taking of the Jews played a central role in the modernization of the states in eighteenth-century Europe. In the Polish-Lithuanian Commonwealth, Prussia and Bohemia, Jews were the first to be so systematically enumerated on a statewide scale. In order to reconstruct fiscal condition and civic order, states made extensive use of censuses of the Jews to control population growth, estimate taxation, and limit residence rights. This direct extension of state power (or “biopower,” to use Foucault’s term) over an individual subject often exceeded the governance capacities of the local officials, who sought optimization at the expense of knowledge. At the same time, Jews actively negotiated their position by hiding and reclaiming control over their status. The paper argues that the lack of ritual meaning of a census in Judaism (in contrast to the Christian tradition) and the routinization of the process on the part of the administration were the key factors that hindered the census taking and, consequently, the effective modernization of states in east-central Europe.

The paper traces the development of censuses between three main contributors, the state, census takers and respondents, by examining the legal context and enumeration practices based on a novel dataset consisting of almost 145 000 individuals living in 53 regions between 1764 and 1809. In tracing the spatial patterns of census taking, standard measures of child sex ratios and child proportions in the census lists of the Jews are compared with the non-Jewish data.

Śławomir Brzezicki • *From Handbooks to the Dehio Alliance. The digital transformation of the “Dehio-Handbuch der Kunstdenkmäler”—from the traditional format of national topographies of art and architecture to an international research infrastructure*

The program for compiling a handbook of German historical monuments (“Dehio-Handbuch der Kunstdenkmäler”), formulated in 1900 by Georg Dehio, a professor of art history at the University of Strasbourg, was intended to provide statistical data on historical monuments in conjunction with the inventories that had been compiled since the late 19th century. It was to be “true to its name: a handy format, easy to transport, with a content structure guaranteeing quick access to information, and just as convenient to use at a desk as it is while traveling.” Subsequent editions of the series—originally covering the German Reich, followed by a separate series for Austria beginning in 1923, and one for Poland starting in 2006—demonstrate that the handbook represents a continuum of a certain standard for recording information about architectural and artistic monuments. New editions and expansions have revised the definitions of cultural landscapes, selection criteria, and time frames, while maintaining the structure of the catalog entries for individual objects that was established from the outset.

With the beginning of a series dedicated to the Baltic republics at the Herder Institute in 2017, and shortly thereafter a retro-digitization project for the current volumes of the German series, the development of a research infrastructure began—the Dehio Alliance (<https://dehio.org>)—aimed not only at digitizing texts, but also to develop a new standard for structured information on historical monuments, offering broader possibilities for statistical searching and linking to other digital knowledge sources. The Linked Open Data environment also requires the use of information to link and co-create authority data for people, corporate bodies, places, and objects, as well as historical gazetteers and multilingual thesauri.

Session 4

Anton Kotenko • *Subscribing to Nationalism: Mapping the Social and Territorial Composition of the Ukrainian National Movement in the Romanov Empire, 1906–1911*

Histories of nationalism often privilege prominent leaders over the broader constituencies of anonymous “activists,” “nationalists,” and “patriots” sustaining national movements. Building on Miroslav Hroch’s study of the formation of modern nations, this article examines the sociospatial composition of the Ukrainian movement between 1906 and 1911 through subscribers to *Council*, the Romanov Empire’s only Ukrainian-language daily, on the basis of Mykola Holoborodko’s

unpublished survey of approximately 5,000 newspaper subscription cards. Treating subscription not only as a definitive evidence of activism, but also as a form of material support, the presentation shows that among subscribers whose occupations Holoborodko could identify, teachers and members of the free professions formed the largest groups; teachers were particularly prominent in villages, where they served as intermediaries between the Ukrainian-language public sphere and rural communities. Spatially, subscriptions were concentrated in Kyiv and Poltava provinces but distributed unevenly across the territory claimed as Ukrainian national space. Between 1906 and 1911, the readership also became increasingly dispersed across the Romanov Empire. The article argues that subscriber records can illuminate the uneven process through which an imagined national space acquired a socially and geographically embedded constituency.

Victoria Gerasimova • *Beyond Coercion and Conviction: A Database Approach to Jewish Conversion in the Kingdom of Poland (1867–1914)*

Jewish conversion in the Russian Empire has long been framed as a product of state pressure — discriminatory legislation, pogroms, forced assimilation. Yet the historical reality in which individual neophytes made their choices was far more complex. How do we capture that complexity without imposing the researcher's own hierarchy of motives? This paper argues that database construction itself becomes a methodological intervention. Drawing on an original dataset of conversion cases from the Kingdom of Poland, the study employs prosopographical analysis to identify behavioral patterns, migration strategies, and social networks that resist reduction to a single explanatory logic. Rather than adjudicating between coercion and agency, the quantitative approach reveals the structural conditions within which choices were made — and exposes how imperial mass records simultaneously document and distort the lives they register.

Edyta Chomentowska, Joanna Kumor-Mielnik • *The Reliability of 19th- and 20th-Century Diocesan Schematisms as a Basis for Reconstruction of the Parish Network*

The reconstruction of the Catholic Church's parish network based on church schematisms (directories) from the 19th and 20th centuries involves entering source data into a database (source layer), followed by the harmonisation of records from different years, partial verification against external sources, and the plotting of results (critical layer) onto maps (GIS). This allows the evolution of church structures to be reconstructed within a changing historical context. At the same time, this context may be one of the factors determining the degree of reliability of the church registers. A source-critical analysis of these censuses demonstrates their high reliability despite interference from censorship, mainly by the

partitioning and communist authorities. The inaccuracies identified are primarily data copied over from previous years. There are also elements resulting from the policies of church authorities (e.g. the inclusion of establishments of disputed status), state authorities (e.g. the removal of formally closed establishments) or social and national ambitions (e.g. the names of parish villages in the Apostolic Administration of Lemkivshchyna). Nevertheless, their reliability is very high, particularly for the interwar period and in the reconstruction of the parish network itself and its selected elements, such as the dedications of churches or their building materials.

Magdalena Baran-Szołtys • *From Archive to Interface: Galicia and the Re-Mapping of Layered Historical Spaces*

While much digital humanities research focuses on the processing and analysis of large-scale historical datasets, this paper intervenes at the level of epistemological structures of spatial and narrative archives, drawing on the theoretical framework developed in my monograph *Galizien als Archiv* (2021, V&R unipress). I conceptualize Galicia as a transnational archive constituted through literary texts, family narratives, material traces, and spatial practices. Within this framework, I distinguish between an “imaginative archive” composed of texts, maps, memories, images, and cultural narratives, and a “spatial archive” anchored in the material post-Galician landscape through buildings, ruins, linguistic traces, culinary practices, or lived experience.

Within this conceptualization, historical regions such as Galicia emerge as spaces shaped by overlapping yet divergent national archives and memory cultures. Although these archives occasionally intersect, they often produce competing narratives, selective memories, and distinct affective mappings of the same places, as exemplified by the multilayered representations of cities such as Lviv/Lwów/Lemberg.

Travelling through historical regions can therefore be understood as a dynamic process of selecting, actualizing, and re-inscribing archival materials in space. Historical space emerges not as a stable geographical entity, but as a continually negotiated relationship between memory, movement, and present-day experience. This raises the question: how do people move affectively, narratively, and culturally through historically layered spaces?

Based on this theoretical model, the paper proposes a participatory digital application for historical regions. What might a historically layered, participatory archive-in-motion look like? Combining geolocation, archival materials, literary quotations, historical maps, and user-generated contributions, such an interface could function as a “living archive” that allows travellers not only to access multiple historical layers of a place, but also to encounter competing narratives and actively contribute to the ongoing production of cultural memory.

Steven Seegel • *Squaring the Triangle: Soil, Sovereignty, and Science in the Belarusian Case of Geographer Arkadz Smolich (1891–1938)*

My article takes a post-positivist discursive approach to key epistemic issues of measurability, the convergence of human geography and data science, and the critical geopolitics of territorial sovereignty through East European geographical discourse. As a case study, I look here at the works of acclaimed Belarusian soil scientist Arkadz Smolich (1891–1938). Smolich was a founding academic expert in geography and one of the 20th century's lost "map men," my category for colonial-era transnational actors who made use of statistical demography, earth sciences, and data processing. Smolich was an advocate for the first independent Belarusian Republic (BNR), and a citizen of the and Belarusian Soviet Socialist Republic (BSSR). During Stalin's purges and 'science wars' of the Soviet 1930s, the social democrat was arrested, transferred into the interior, and executed in 1938 by the NKVD. My case study takes a fresh look at the general silence about Smolich's life and scholarly reputation into the Cold War era followed by his 'heroic' revival in Belarusian nationalizing circles of the 1990s and 2000s.

1 October, IH PAN

Session 5

Gábor Demeter • *Integrating Historic Databases in Hungary: the GIS_{ta} Hungarorum (1720–1910), the RD1500 (1330–1520) and the Collection of Ottoman Registers (1546–1590)*

The GIS_{ta} Hungarorum project began in 2014 with the goal to organize and integrate settlement-level census and tax conscription data of the Kingdom of Hungary derived from serial sources between the 18th and 20th centuries into an extendable database for the long-term analysis of regional patterns in socio-economic and demographic processes. From the beginning we also aimed to create a framework that allows the integration of medieval and 21st-century databases under construction. By creating indicators from the collected raw variables it becomes possible to measure socio-economic advance and regional patterns of inequalities for 5 time horizons ([1720](#), [1786](#), [1828](#), [1880](#), 1910) offering a new tool for historians and geographers. The creation of a user-friendly platform for public use that allows the sharing of data and results on thematic maps was also among the aims. In 2026 Croatian database (GIS_tatistica Croatica, [1786](#), [1828](#), [1910](#)) was added to the application. An ESDA-style processing of settlement-level allows the

- (1) the examination of the spatial patterns of socio-economic and demographic phenomena, the identification of the nature and extent of inequalities, and the reconstruction of rural development conditions at several time horizons ([1720](#), [1780](#), [1828](#), [1880](#), 1910) (static picture).
 - (2) the examination of the spatial patterns of socio-economic-demographic conditions, inequalities, and changes in development conditions over time (process dynamics).
 - (3) the visualization and evaluation of the above mentioned phenomena through maps and also historical statistical methods.
 - (4) the determination of historical regions (clusters) based on socio-economic and demographic indicators and the analysis of their compactness/fragmentation and distinctive features
 - (5) the analysis of socio-economic and demographic indicators for non-quantifiable categories (religion, language, land ownership-types, legal status of settlements).
 - (6) In addition to historical statistical data, external variables (soil conditions and topography considered conservative, availability of raw materials, distance from water sources and roads, etc.) were also integrated into the database.
 - (7) In order to move towards applied science and practical use, the idea to link the historical database with databases related to contemporary, modern research on peripheries has also emerged as a goal.
-

Ruth Mostern • *Naming Places: Gazetteers and State Power from Print to Pixels*

Historically, a gazetteer is a dictionary of named places, typically listed in alphabetical order. Though different gazetteers may emphasize particular concerns, the entries for each place named in a gazetteer generally include information like its location (according to coordinates, proximity to waterways and other natural features, or containment in a larger administrative unit), its population, aspects of its culture and society, and characteristic natural or manufactured products associated with it. This talk explores gazetteers and state power from two angles.

First, it is a global history of gazetteers demonstrating that the transition to encyclopedic gazetteers from earlier forms of spatial recordkeeping accompanied rising ambitions for knowledge management by governments that sought to expand territorial power through military control and tax collection. Gazetteers were essential tools in this process as named places came to be linked to many state functions. Gazetteers and maps played different though complementary roles as spatial knowledge organization systems during these centuries.

Second, this talk focuses on the process of converting historical print gazetteers into digital information. I will demonstrate the workflow that the World Historical Gazetteer (WHG) has used to digitize and map the 1856 *Gazetteer of the World* and the lessons we are learning from that effort. I will also introduce other datasets from the World Historical Gazetteer that shed light on the history of

state power. Finally, I will promote the WHG Linked Open Data methodology and the WHG approach to linked attestations about places as solutions for accommodating contested, ambiguous, multilingual, and frequently changing attributes that are always a characteristic of historical places and contended state power, very notably in East Central Europe.

Kristina Wittkamp • *Archives Reconnected: A FAIR Hub for Eastern Europe's Digitized Heritage: A Hypothèses Blog Bridging Archives, Research & Education* (<https://archiveome.hypotheses.org/>)

This above linked newly launched *hypothèses* blog project bridges critical gaps in access to digitized archives across Eastern and Central Europe through a FAIR-compliant database (Findable, Accessible, Interoperable, Reusable). It systematically links metadata from approx. 130 archival sources in 14 countries (e.g., Belarus, Ukraine, Poland, Lithuania, Estonia, Latvia, Slovenia, Serbia, Bosnia and Herzegovina, Romania, Georgia, the Czech Republic, Bulgaria and Croatia), excluding Russia, offering the first centralized resource for these underrepresented materials.

Developed in three phases (collection, Excel structuring, public launch), the blog database organizes records by region, period, and topic, enabling researchers—especially early-career scholars—to navigate fragmented and vulnerable digital holdings. Future expansions will integrate datasets supported by the University of Passau's Research Data team for DH compatibility.

The blog project safeguards endangered cultural heritage (e.g., wartime records, dissident archives) in politically volatile regions while fostering transnational dialogue. It also serves as a dynamic teaching tool: its materials are deployable in digital history curricula, training students in metadata standards and critical analysis.

By democratizing access to marginalized archives, the project models a scalable solution for preserving contested heritage. Its dual focus—combining preservation with pedagogy—creates a feedback loop: users enrich its database while engaging with once-inaccessible sources. This initiative not only mitigates cultural memory loss but also empowers a new generation of researchers to reshape scholarship on Eastern Europe.

Adrianna Sznepik, Piotr Jaskulski • *Interpreting Mass Historical Sources with LLMs: The Case of The Geographical Dictionary of the Kingdom of Poland*

Large language models (LLMs) are increasingly used in historical research as tools supporting tasks such as search, classification, and information extraction. One particularly promising application is the transformation of unstructured textual sources into structured data.

This paper presents a case study of the *Geographical Dictionary of the Kingdom of Poland* (1880–1902), processed within the Vademecum project. As a first step, a preliminary survey of entries relating to localities (ca. 150,000) was conducted by historians, who identified 57 thematic categories, including administrative and ecclesiastical affiliation, statistical data, economic and social structures, and cultural heritage. LLMs were then used to extract and assign information to these categories. However, verification of the resulting structured data revealed a number of challenges, related not only to accuracy but also to ambiguity and historically specific categories embedded in the source. We argue that the use of LLMs introduces a new layer of interpretation between the source and the historian. This raises methodological and epistemological questions: where is the boundary between data extraction and interpretation? How does the use of AI affect source criticism? And what is the role of the historian in a workflow increasingly mediated by generative models?

Tadeusz Epsztein, Michał Raczkowski • *Large Landownership at the Turn of the Nineteenth and Twentieth Centuries: Socio-Economic History and AI and LLMs*

One of the key research problems in the history of large landownership at the turn of the nineteenth and twentieth centuries is determining the scale of landholdings and their spatial distribution. This article addresses this issue on the basis of mass sources, such as the *Geographical Dictionary of the Kingdom of Poland and Other Slavic Countries* as well as statistical compilations concerning the lands of the Prussian Partition, Galicia, and selected governorates of the Russian Empire. These sources are characterised by considerable heterogeneity—both in terms of format (narrative text versus tabular data) and in the categories and units of measurement employed.

The paper proposes the use of artificial intelligence tools, in particular large language models (LLMs), for the automatic extraction of data and the creation of a unified database of information on large landownership. This makes it possible to conduct quantitative and spatial analyses on a scale not previously attainable. At the same time, the paper reflects on the limitations of this approach, pointing to the need for a critical interpretation of historical data and into account their epistemological and political context.

Session 6

Elżbieta Nowosielska • *Newspaper Printing in the Prussian Partition between 1865 and 1918*

The territories of the former Polish-Lithuanian Commonwealth incorporated into Prussia became a site of intensive Germanisation, which often produced effects opposite to those intended. The press

published by local ethnic groups functioned both as a tool of Germanisation and of resistance. Although Polish publications were strictly controlled and their editors frequently faced serious consequences, the press continued to develop, supported by new printing techniques and the growth of mass culture.

This presentation aims to examine the network of printing houses producing periodicals, and calendars in the Prussian Partition. It is based on material from fasciculi 2, volume 3 of the “Bibliografia Historii Polski XIX i XX w.”, currently being revised within the project “The Cultural and Intellectual Geography of the Former Polish Lands after the Partitions, 1865–1918: A Digital Vademecum”.

The study raises questions about the relationship between ethnic structure and printing activity, differences between Polish- and German-owned printing houses, the existence of multi-ethnic services, and the influence of state policy on the development of printing centres.

Victoria Gerasimova • *Mapping the Unmapped: Jewish Presence in the SGKP and the Limits of Cultural Heritage GIS*

The Słownik geograficzny Królestwa Polskiego forms the textual backbone of the Vademecum GIS project reconstructing the cultural landscape of former Polish territories. Yet this gazetteer, which remains to this day one of the most important sources for the historical and cultural geography of the Polish lands, was compiled within a particular national and confessional framework: how does it represent ethno-confessional minorities — and what does this mean for the data we map? This paper uses computational methods (corpus statistics, collocation analysis, topic modelling, word embeddings) to examine the representation of Jews, the region’s most visible minority. I argue that the SGKP encodes Jews primarily as demographic and economic data points rather than cultural agents — a framing reproduced and amplified when the text becomes GIS infrastructure — illuminating both the possibilities and structural blind spots of heritage mapping built on nineteenth-century national gazetteers.

Wiktor Werner • *Measuring Authoritarian Environmental Governance: Interpretable Machine Learning and the Historical Analysis of SB Activities in Late Socialist Poland*

This paper explores the epistemic value of interpretable machine-learning methods for the analysis of twentieth-century mass historical data. It draws on 58 archival case files concerning water pollution investigated by the *Służba Bezpieczeństwa* (SB) in socialist Poland between 1980 and 1989 and combines source criticism with decision trees, random forest, logistic regression, and clustering.

Its central theoretical claim is that, in digital history, classical machine-learning algorithms may be as scientifically significant as more complex neural architectures, precisely because they remain more transparent, controllable, and open to historical interpretation. In the case examined here, the main epistemic gain does not lie in predictive power alone, but in the ability to identify and interpret the institutional variables most strongly associated with modernisation, including referral to the prosecutor's office, preventive measures, and the type of responsible actor.

The paper argues that interpretable models are particularly well suited to historical inquiry because they preserve a connection between digital processing and hermeneutic accountability. Rather than treating black-box complexity as the highest methodological standard, it proposes controlled interpretability as a major epistemic virtue in the analysis of modern bureaucratic archives from East Central Europe.

Timothy Tambassi • *No Neutral Data: A Perspectivist View of Historical Sources*

This talk examines the epistemological and methodological implications of applying geographical perspectivism to the digital processing of geo-historical mass sources. I argue that such sources cannot be treated as neutral repositories of facts. Rather, following the main claims of geographical perspectivism, according to which each geographical perspective reflects specific cognitive partitions, aims, and cultural assumptions, I propose that digital geo-historical databases should accommodate categorial pluralism and potential inconsistencies rather than enforce a single consistent ontology. This approach allows historians to preserve the epistemic richness of source-creators' worldviews while using digital tools like AI for information extraction. Ultimately, the talk advocates for an open categorial framework in geo-historical data editing—one that balances technical standardisation with critical sensitivity to how past actors categorised space, identity, and belonging.



Ministerstwo Nauki
i Szkolnictwa Wyższego

