András Simon[1]
Ildikó Szerényi[2]

AI-POWERED DATA ENRICHMENT: REVITALIZING GEOGRAPHICAL DATA AT THE NATIONAL ARCHIVES OF HUNGARY

Abstract

Purpose: *The purpose of this paper is to provide a conceptual and practical overview of namespace development, with a particular focus on the creation and renewal of the geographical namespace at the National Archives of Hungary.*

Methodology: *The study combines theoretical analysis with a case study approach. It reviews the structure and function of namespaces in archival systems and analyzes the process of building a local geographical namespace. The methodology includes data modeling, authority record enrichment, and the integration of automatic, AI-supported matching techniques for disambiguating place names.*

Results*: The paper outlines the practical steps and challenges involved in creating a comprehensive geographical namespace. It presents a model for linking variant forms of place names to unique identifiers and demonstrates how enriched authority records improve search accuracy. The results also highlight the potential for future integration with global name authorities.*

Conclusion: *Namespaces significantly enhance the precision, consistency, and interoperability of archival metadata. The case study of the National Archives of Hungary demonstrates that namespace development requires both conceptual clarity and technical infrastructure. The findings suggest that well-structured namespaces, especially when aligned with international standards, offer substantial benefits for both internal data management and external discoverability.*

Key words: *metadata quality, authority control, namespace, geographical names, memory institutions, archival description, interoperability*

1 András Simon, IT developer at the National Archives of Hungary, Digital Services Department, e-mail: simon.andras@mnl.gov.hu.
2 Ildikó Szerényi, senior archivist at the National Archives of Hungary, e-mail: szerenyi.ildiko@mnl.gov.hu.

ARRICCHIMENTO DEI DATI BASATO SULL'INTELLIGENZA ARTIFICIALE: RIVITALIZZAZIONE DEI DATI GEOGRAFICI PRESSO L'ARCHIVIO NAZIONALE UNGHERESE

Abstract

Scopo: *lo scopo di questo articolo è fornire una panoramica concettuale e pratica dello sviluppo dei namespace, con particolare attenzione alla creazione e al rinnovamento del namespace geografico presso l'Archivio Nazionale Ungherese.*

Metodologia: *lo studio combina l'analisi teorica con un approccio basato su casi di studio. Esamina la struttura e la funzione degli spazi dei nomi nei sistemi di archiviazione e analizza il processo di costruzione di uno spazio dei nomi geografici locale. La metodologia include la modellazione dei dati, l'arricchimento dei record di autorità e l'integrazione di tecniche di corrispondenza automatiche supportate dall'intelligenza artificiale per la disambiguazione dei nomi di luoghi.*

Risultati: *Il documento delinea i passaggi pratici e le sfide coinvolte nella creazione di uno spazio dei nomi geografico completo. Presenta un modello per collegare forme varianti di nomi di luoghi a identificatori univoci e dimostra come i record di autorità arricchiti migliorino l'accuratezza della ricerca. I risultati evidenziano anche il potenziale per una futura integrazione con le autorità di nomi globali.*

Conclusioni: *I namespace migliorano significativamente la precisione, la coerenza e l'interoperabilità dei metadati archivistici. Il caso di studio dell'Archivio Nazionale Ungherese dimostra che lo sviluppo dei namespace richiede sia chiarezza concettuale che infrastrutture tecniche. I risultati suggeriscono che namespace ben strutturati, specialmente se allineati agli standard internazionali, offrono vantaggi sostanziali sia per la gestione interna dei dati che per la reperibilità esterna.*

Parole chiave: *qualità dei metadati, controllo dell'autorità, spazio dei nomi, nomi geografici, istituzioni della memoria, descrizione archivistica, interoperabilità*

OPLEMENITENJE PODATKOV Z UMETNO INTELIGENCO: REVITALIZACIJA GEOGRAFSKIH PODATKOV V MADŽARSKEM NARODNEM ARHIVU

Izvleček

Namen: *Namen prispevka je podati konceptualni in praktični pregled razvoja imenskih prostorov, s posebnim poudarkom na oblikovanju in prenovi geografskega imenskega prostora v Madžarskem narodnem arhivu.*

Metodologija: *Raziskava združuje teoretično analizo s študijo primera. Pregleduje strukturo in funkcijo imenskih prostorov v arhivskih sistemih ter analizira proces oblikovanja lokalnega geografskega imenskega prostora. Metodologija vključuje modeliranje podatkov, oplemenitenje normativnih zapisov ter vključevanje samodejnih, z umetno inteligenco podprtih tehnik ujemanja za razločevanje krajevnih imen.*

Rezultati: *Prispevek predstavi praktične korake in izzive pri oblikovanju celovitega geografskega imenskega prostora. Predstavi model povezovanja različnih oblik krajevnih imen z enoličnimi identifikatorji ter pokaže, kako oplemeniteni normativni zapisi izboljšujejo natančnost iskanja. Rezultati prav tako izpostavljajo potencial prihodnje integracije z globalnimi imenskimi normativnimi bazami.*

Zaključki: *Imenski prostori bistveno izboljšujejo natančnost, doslednost in interoperabilnost arhivskih metapodatkov. Študija primera Madžarskega narodnega arhiva dokazuje, da razvoj imenskih prostorov zahteva tako konceptualno jasnost kot tehnično infrastrukturo. Ugotovitve nakazujejo, da dobro strukturirani imenski prostori, še posebej kadar so usklajeni z mednarodnimi standardi, prinašajo pomembne koristi tako za notranje upravljanje podatkov kot za zunanjo dostopnost.*

Ključne besede: *kakovost metapodatkov, normativna kontrola, imenski prostor, geografska imena, spominske ustanove, arhivski opis, interoperabilnost*

1. INTRODUCTION

There is growing interest in the online services of the National Archives of Hungary, which makes it necessary for users to access archival documents through search interfaces without the assistance of archivists. Readers should be supported by the advanced search features of the web-based catalogues.

One of the most important search parameters for archival documents is the geographical name; therefore, searching for geographical concepts (entities) should be supported by the National Archives' geographical namespace. However, the finding aid databases of the archive contain millions of records and geographical terms. Likewise, the global geographical namespace itself includes millions of name entries.

Geographical names often appear in multiple forms within the finding aid records, and a single term can refer to several distinct geographical concepts. In the previous system, matching was performed using human-defined rules. As the volume and diversity of data increased, however, this rule-based approach alone proved insufficient, since small changes in data formats or structure require manual updates and writing and maintaining hundreds or thousands of rules is labor-intensive. Consequently, it became necessary to introduce an automated, machine learning–based system capable of recognizing new patterns and relationships without the need for manually defined rules for each individual case.

In addition to the similarity of individual geographical names, other informational elements are also considered when matching the name forms found in archival documents with the corresponding geographical concept in the namespace. This paper describes the AI-based solution developed for this task, providing an overview of the application's model construction and training process.

Over the past three decades, memory institutions have increasingly adopted integrated collection management systems built on relational databases. The data elements within these databases are typically published on the Semantic Web, as their core units — known as *entities* — are defined using persistent identifiers. Persistent identifiers (such as URIs) ensure that each entity — be it a person, a place, or an organization — can be uniquely and unambiguously referenced across different systems and over time.

The Semantic Web extends the traditional World Wide Web by enabling machines to understand and process the meaning — or semantics — of data. While the traditional web primarily presents information in formats readable by humans (like text or images), the Semantic Web structures data in a way that allows computers to interpret relationships between entities. This makes it possible, for example, to link a historical figure mentioned in an archive to their biography in a digital encyclopaedia, or to unify variant place names across multiple datasets. The identifier string of an entity, together with its associated name variations, constitutes a *namespace record*. This record not only captures the preferred label for an entity but also includes variant names, language-specific forms, dates, hierarchical relationships, and other descriptive attributes. These enriched namespace records provide both human-readable and machine-readable access to the entity's identity and context.

Memory institutions most commonly create namespaces for personal names, corporate bodies, events, geographical locations, and common (subject) terms.

These namespaces help ensure consistency in cataloguing and enhance interoperability between institutions. When published on the Semantic Web, such namespaces become powerful tools for linking datasets, enabling cross-institutional discovery, and supporting digital humanities research.

Some libraries and museums are developing namespaces within their own collection management systems, based on their existing authority records. In Hungary, the most significant and well-structured namespaces have been created by institutions with the largest and most diverse collections, such as the Petőfi Museum of Literature (PIM), the National Széchényi Library, the Library of the Hungarian Academy of Sciences, the Szabó Ervin Library of Budapest, and several major university libraries.

In addition to building local namespaces, many Hungarian institutions — especially museums — have begun integrating external, globally recognized vocabularies. These include the GeoNames open-source geographical database, which covers all countries and contains over eleven million place names, and the Getty Research Institute's Art & Architecture Thesaurus (AAT), a controlled vocabulary used to describe items of art, architecture, and material culture. These external resources support standardization and international interoperability, allowing for richer metadata and broader discoverability.

At the National Archives of Hungary (NAH), the development of institutional namespaces has been underway for more than a decade.

To support the creation of finding aids (catalogues of archival documents), the NAH's namespace database includes structured records for personal names, corporate entities, and geographical locations. These namespaces are extensively used in daily archival work, but they are now considered ready for conceptual and functional revision, as well as significant content expansion.

Given the central role of place names in archival description and retrieval, the decision was made to begin the renewal process with the geographical namespace. This namespace will serve as a pilot for rethinking data models, authority structures, and technical frameworks. Once the renewed geographical namespace is complete, the process will continue with the personal and corporate name authorities, applying the lessons learned and the model refined during the initial phase.

An important aspect of the project was implementing an existing international data model in the database and aligning the migrated namespace elements with this structure. The international data model used is based on a framework developed and published by the Getty Research Institute specifically for building namespaces.

This paper outlines both the guiding principles and the key stages of the geographical namespace project at the National Archives of Hungary. It details the migration process from various data sources into a unified namespace. Although the source data is often redundant and heterogeneous, the project preserves the original data elements unchanged. Where multiple records refer to the same geographical concept, they are linked and enriched with geographic coordinates to improve accuracy and semantic clarity.

Simple string comparison of geographical names proved inadequate for identifying and consolidating matching concepts. To overcome this limitation, an AI-based algorithm was developed to support the process. This algorithm was also used to enrich the geographical terms found in the metadata of finding aid records by linking them to corresponding namespace entities.

The paper also discusses how the enhanced namespace has been integrated into the online public services of the National Archives of Hungary, improving both searchability and user experience through more accurate and semantically linked metadata (Ungváry, 2012).

2. THE CONCEPT OF NAMESPACE

The development of national namespaces began nearly two decades ago. The reason for their creation lies in the vast number of proper and common names found in digital texts across the web. In this enormous pool of names, traditional alphabetical searches in dictionaries prove inadequate; therefore, the use of semantically structured namespaces has become necessary.

The development of Hungarian namespaces also plays an important cultural role by supporting the use of the Hungarian language beyond the country's borders.

When discussing namespaces, it is important to define their classification.

Kinds of Namespaces According to Area:

- International
- National
- Local
- Thematic

Kinds of Namespaces According to Namespace Unit:

- Persons
- Corporate Bodies
- Geographical Names
- Events
- Concepts

Namespaces are extensive, well-structured datasets designed to organize and manage information consistently. Each element within a namespace is validated and linked to other elements through clearly defined relationships and directions. This ensures that data is not only accurate but also interconnected in a meaningful and machine-readable way. Due to these features, memory institutions — such as archives, libraries, and museums — often develop multiple namespaces based on their structured databases.

Institutional databases — such as library catalogues, museum inventory books, and archival finding aids — rely heavily on controlled vocabularies or standardized lists of values. These vocabularies typically include various types of names and concepts: personal names (e.g., authors, historical figures), corporate names (e.g., organizations, government bodies), geographical names (e.g., cities, regions), as well as zoological, botanical, and geological species, and abstract

concepts. Such vocabularies ensure consistency in data entry, facilitate search and retrieval, and provide a foundation for building interoperable namespaces.

At the National Archives of Hungary, such structured vocabulary elements are also recorded within the metadata entries of archival finding aids. These digital metadata records describe both analogue and digitized archival documents, containing key details such as creators, locations, dates, and subjects. Many of these data elements are suitable for reuse as part of our namespaces.

We are currently developing dedicated personal and geographical namespaces based on these records. Additionally, we plan to create a corporate namespace to represent organizations and institutions found within our archival holdings. The aim is to improve metadata quality, enhance searchability, and support semantic linking across our archival systems.

PERSONAL AND GEOGRAPHICAL NAMESPACES

Workflow for Creating a Personal or Geographical Namespace Record

1. Identifying a personal or geographical name in an archival document
2. Entering name-related information into the finding aid metadata
3. Integrating personal or geographical names into controlled lists of values
4. Creating authority records
5. Generating namespace records

The process begins when an archivist identifies the name of a person or a geographical entity in an archival document. This name is then entered into the digital metadata of the finding aid database. However, recording a name in the database is not limited to simply typing a string — it also involves enriching the controlled vocabulary of personal or geographical names.

Once identified and validated, these names are formalized as authority records, which serve as the building blocks of a namespace. By establishing hierarchical or associative relationships among them, a semantically structured namespace can be constructed (D'Souza, 2025).

IDENTIFICATION OF GEOGRAPHICAL RECORDS – KEY CONSIDERATIONS:

Type of Name: The name may be a unique proper noun or a commonly used descriptive term (e.g., Greenwich vs. Springfield) "Greenwich" refers specifically to

a district in London known for the Prime Meridian and its maritime history. The name is distinct and not commonly used elsewhere, making it easy to identify. "Springfield" is used by many towns and cities across the United States (over 30 in total), so it requires additional context (such as state or country) to disambiguate. It's not unique and could refer to multiple places.

Name Length: Geographical names may vary significantly in length, from very short (e.g., Soorn) to unusually long (e.g., Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch) the longest place name in Europe.

Spelling Variations and Errors: Data sources may contain alternate spellings (e.g., Český Krumlov vs. Böhmisch Krumau). Both refer to the same town in the Czech Republic. The variation reflects historical and linguistic context — Czech vs. Austrian/German usage. Simple misspellings: "Yokohama" (correct) and "Yokohoma" (misspelled), a simple transposition of letters could prevent accurate matching in namespaces.

Historical Name Changes: Place names may change over time due to linguistic reforms or geopolitical shifts (e.g., Aarhus vs. Århus; Lemberg (Austro-Hungarian Empire), Lwów (Poland), Lvov (Soviet Union), Lviv (Ukraine)) All these names refer to the same city, now in western Ukraine. Recognizing these variants is essential for historical research, genealogy, and data reconciliation.

Type of Geographical Entity (Role): The same name may refer to different types of entities — such as a settlement, river, or hill — or may appear in multiple languages with similar meanings (e.g., peak, pinnacle)

Geographical Coordinates: Coordinates may be missing or inaccurate in the source data, which complicates precise identification.

Hierarchical Relationships: A geographical place may be a subpart of a broader geographical concept (e.g., a village within a municipality)

Thesaurus Forms: Names may appear either as preferred terms or variants in controlled vocabularies and thesauri. The preferred term is the standardized, authoritative version of a name used for indexing and cataloging, ensuring consistency across systems and institutions. Variant terms, on the other hand, include alternative spellings, historical forms, translations, local names, or common abbreviations. These variants are linked to the preferred form.

3. BUILDING A GEOGRAPHICAL NAMESPACE BY MERGING FOUR DATA SOURCES

Following the identification and extraction of various personal and geographical name entries, hypothetical records representing persons and geographical entities are created within the database environment. While the personal namespace of the National Archives of Hungary (NAH) is built using name elements found in the metadata of finding aids and full-text archival documents, the construction of the geographical namespace relies on the integration of four major data sources:

- MNL GEO – The internal geographical database of the National Archives of Hungary containing 70,849 records.
- Geotaurusz – The Hungarian thesaurus of geographical concepts, consisting of 109,008 entries.
- GeoNames – A global geographical database featuring 12,237,573 records.
- The Digital Atlas of Medieval Hungary *(compiled by Pál Engel)* – A specialized historical geographical database with 75,031 records.

By harmonizing and linking these sources, the resulting namespace offers a robust and enriched framework for identification, organization, and retrieval of geographical information within the archival context (Bánki et al, 2023).

3.1 BASIC PRINCIPLES AND STEPS FOR BUILDING A NAMESPACE

(Creating a database of entities and enriching data by linking them)

- Final goal: One concept → one database entity, derived from a single authoritative data source, with all variant names, spellings, historical forms, and language versions linked to that single, unified representation to ensure consistency, disambiguation, and interoperability across systems.
- Term-entity relationship: A single database entity can be linked to multiple terms (i.e., name variants)
- Term ambiguity: A string may function as the preferred term for one concept and as a variant for another (Neuhaus – Selbitz / Jindřichův Hradec). In this example Neuhaus refers both to a German town and as the historical German name for Jindřichův Hradec.
- Rule based connections: Entities are linked based on identical strings (name and role) and identical or nearly identical geographical coordinates

- AI based model connections: The system identifies potential matches by measuring how similar two names are using string comparison methods (like Soundex or Jaccard), and by checking whether their geographic coordinates are close enough to suggest they refer to the same place.
- Training datasets: Carefully chosen, predefined datasets are used for training the AI model and for human validation. The size and careful selection of these datasets are critical to the success of the model.
- AI learning process: The AI application generates its own internal rules for predicting relationships between entities
- Model refinement: Based on validation results, the AI model can be further developed and fine-tuned.

In following these steps, we utilize an AI-based application. To illustrate the rationale behind using AI for building a namespace database and enriching data, we have prepared a SWOT analysis. A SWOT analysis is a strategic planning tool used to evaluate a project, organization, or idea by examining four key aspects: Strengths, Weaknesses, Opportunities, Threats.

Strengths: By analyzing datasets with appropriate software, we can analyze data and get good results. Different types of information needed for decisions cannot be easily compared using traditional engineering methods, but AI-based software can build models with guidance from human validation	Weaknesses: Such a quantity of data cannot be appropriately handled by a human without creating hypotheses and preconceptions, which could mislead us. Selecting improper data elements for training and validation can result in inaccurate data models.
Opportunities: Huge amounts of data are available from archival documents. Concepts (personal, geographical and corporate units) can be easily defined. Namespaces developed by other projects may be utilized.	Threats: Lack of information in both archival documents and the data sources of the namespace. Similar strings for different names and different strings for the same data entities

Example of a namespace element: Braşov (Romanian) – Brassó (Hungarian) – Kronstadt (German)

Braşov was part of the Kingdom of Hungary until the end of the World War I. Since 1918, it has been part of Romania. The town had a significant German (Saxon) population with great autonomy until the end of the 19th century. As a result, its Hungarian, Romanian and German names are well-known and frequently appear in various documents.

In Table 1, we demonstrate the various occurrences of this town in the namespace. The table columns are as follows:

Entity_id: The identification number of the geographical concept (the town) within the namespace.

Subject_id: The identification number of the geographical concept in the original data source.

Term_id: The identification number of the specific data string in the original data source.

P/V: A letter indicating whether the term is a Preferred version (P) or a Variant (V). Usually the term_id is the same as the subject_id when the term is preferred, but this is only for technical reasons and does not always apply.

Latitude: Geographical latitude coordinate of the geographical object.

Longitude: Geographical longitude coordinate of the geographical object.

Data source: Name of the data source.

Table 1: Entries for the town of Braşov in the namespace of the National Archives of Hungary

Entity_id	Subject_id	Term_id	Term	P/V	Latitude	Longitude	Data source
12331699	8578	8578	Brassó	P			MNL GEO
12331699	8578	8576	Braşov	V			MNL GEO
3164156	3164156	5605904	Kronstadt	P	61,1837	28,24582	Geonames
3173529	3173529	5617550	Kronstadt	P	61,35389	27,41406	Geonames
8374913	8374913	14270133	Braşov	P	45,75	25,33333	Geonames
12331699	8374914	14270138	Braşov	P	45,64861	25,60613	Geonames
12331699	8374914	14270144	Brasov	V	45,64861	25,60613	Geonames
12331699	8374914	14270146	Brassó	V	45,64861	25,60613	Geonames
12331699	8374914	14270150	Kronstadt	V	45,64861	25,60613	Geonames
8485706	8485706	14542899	Kronstadt	P	59,99541	29,76668	Geonames
8488615	8488615	14551213	Ostrov Kotlin	P	60,00598	29,73834	Geonames
8488615	8488615	14551225	Kronstadt	V	60,00598	29,73834	Geonames
9579011	9579011	16857871	Stara Mykolayivka	P	48,32804	37,68157	Geonames
9579011	9579011	16857879	Kronstadt	V	48,32804	37,68157	Geonames
12288489	12288489	20962255	Kronstad Estate	P	-19,6333	32,8	Geonames
12288489	12288489	20962256	Kronstadt	V	-19,6333	32,8	Geonames
12331699	12331699	21020923	Brassó	P			Geotaurusz
12331699	12331699	21034556	Braşov	V			Geotaurusz
12331699	12439044	21178647	Brassó	P	45,6492	25,606754	Engel

4. RESULTS: USING GEOGRAPHICAL CONCEPTS FROM OUR NAMESPACE TO ENRICH THE METADATA RECORDS OF THE FINDING AID DATABASES

The namespace records can be linked to geographical terms of the electronic finding aid database of archival documents. This linking can be performed manually by archivists or automatically by an AI-based software application. The method and workflow of the AI-based application are the same as demonstrated in the previous chapter. By linking a metadata record of an archival document to a geographical namespace unit, the record can be visualized on a map using the coordinates of the geographical concept identified in the document.

4.1 EXAMPLE OF LINKING ARCHIVAL DOCUMENTS TO A GEOGRAPHICAL NAMESPACE RECORD: DRAVOGRAD / DRAUBURG / UNTERDRAUBURG

Dravograd / Unterdrauburg is the name of the same city in Slovenian and German. The city is currently located in Slovenia.

Drauburg (missing from the namespace, it never officially existed)

During the Austro-Hungarian Empire (Österreich-Ungarn), the town was called Unterdrauburg, which is still seen on road signs in Austria today. During the Austro-Hungarian Monarchy, there were two cities on the banks of the Drava River, Oberdrauburg and Unterdrauburg. According to the Treaty of Saint-Germain (1919), Oberdrauburg remained in Austria (in Carinthia), while Unterdrauburg was assigned to Yugoslavia and renamed Dravograd. After the first dissolution of Yugoslavia, the city became part of Germany, and following the second dissolution, Dravograd became part of Slovenia.

The name “Drauburg” refers to a city that never actually existed. Between the two world wars, a Hungarian officer on an official document recorded this name, possibly having spoken with a Yugoslavian person in German and simply translating “Dravograd” into German. As a result, this name format appears in the metadata records of the finding aids but is not present in the namespace. The AI-based linking application correctly linked this term to the concept Dravograd/Unterdrauburg.

In Figure 1, we show the Point of Interest (POI) from the archival document displayed on the map, positioned according to the coordinates from the namespace.

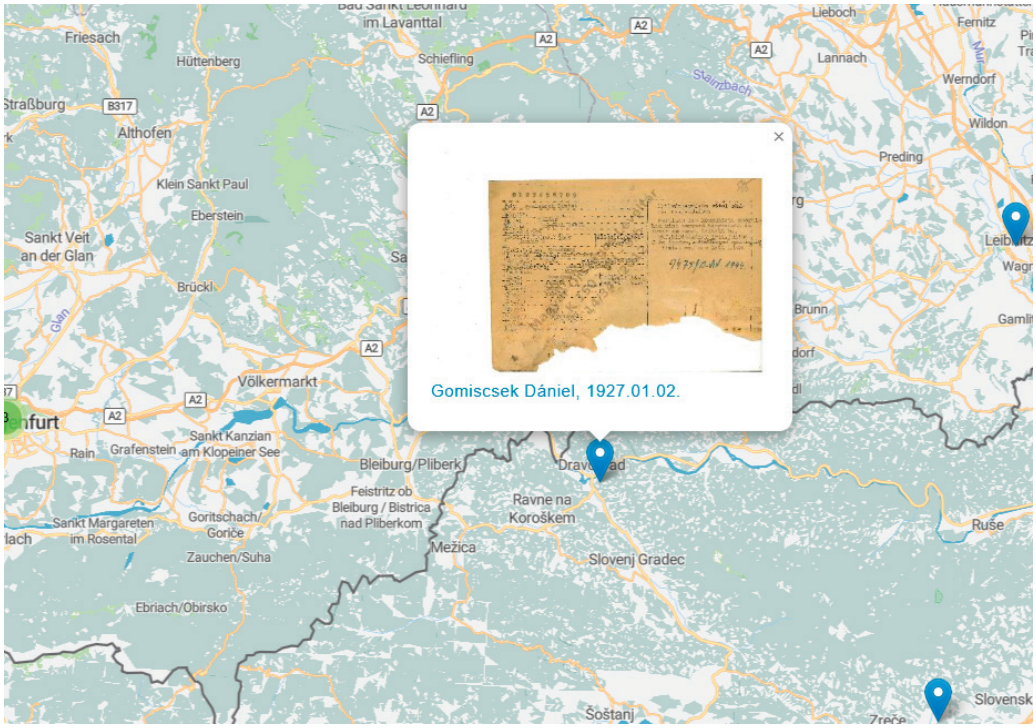

Figure 1: Dravograd on the map, showing the archival document containing the town's name

In Figure 2, we show the metadata record of the document alongside its facsimile. The facsimile can be zoomed in on the public website.

Gomiscsek Dániel, 1927.01.02.

HU-MNL-OL-K-150-VII-9/b 1927.01.02

Név	**Gomiscsek Dániel**
Születési hely	**Drauburg**
Születési dátum	1927.01.02.
Anyja neve	Horváth Cecil
Apja neve	Gomiscsek Izidor
Lakhely	**Muraszombat** - Muraszombat [0.58] - Muraszombat (környék) [0.46] - Muraszombat [0.46]
Vallás	római katolikus
Foglalkozása	tanuló
Ügyleírás	magyar partizán
Képek	

Doboz_03/KMBT_C20020101111153 65
8-6

Figure 2: The metadata record of the document containing the town name Drauburg, along with the digitized document.

In Figure 3 we show the namespace record for Dravograd / Unterdauburg

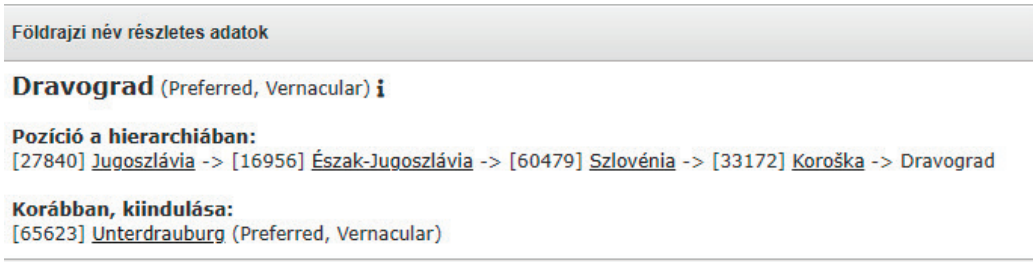

Figure 3: Namespace record of Dravograd/Unterdrauburg

5. CONCLUSION

Enriching the metadata records of archival documents improves the efficiency of searches in the finding aid databases and enables the visualization of documents on a map. We plan to publish all results of geographical data enrichment in archival documents, both as map displays and within archival catalogue records.

We also look forward to building personal and corporate namespaces. Currently we are building the personal namespace of the National Archives of Hungary. After defining the basic principles for the identification of personal records, our current task is to identify individuals based on personal name elements extracted from digitized civil registers. Although a person as an entity is easier to define than a geographical or corporate concept, the large number of similar personal names and the diversity of name variations from digital archival documents make building of a personal namespace challenging. We hope that with the support of AI-based applications, we can accurately identify individuals in our database and share our successful results in the near future.

We would like to thank Zoltán Szatucsek, Zsolt Bánki and Zsolt Záros, colleagues at the National Archives of Hungary, for their valuable contributions to the broader project on which this article is based.

REFERENCES

Bánki, Zs., Szatucsek, Z., & Záros, Zs. (2023). A névterek mint a hiteles tudás forrásai: A Nemzeti Levéltár földrajzi névtér projektjének bemutatása. *Tudományos és Műszaki Tájékoztatás, 70*(4), 472–482. Retrieved at https://doi.org/10.3311/tmt.13273 (accessed on 15.8.2025).

D'Souza, J. (14. 4. 2025). Taming the generative AI Wild West: Integrating knowledge graphs in digital library systems. *Code4Lib Journal*, *60*. Retrieved at https://journal.code4lib.org/articles/18277 journal.code4lib.org (accessed on).

Ungváry, R. (2012). A névtér mint kulturális szükséglet: A nemzeti aggregátor kitüntetett feladata. *Tudományos és Műszaki Tájékoztatás, 59*(8).

Haslhofer, B., Isaac, A., & Simon, R. (2018). Knowledge graphs in the libraries and digital humanities domain. In E. Méndez, F. Crestani, C. Ribeiro, G. David, J. C. Lopes (eds.), *Digital Libraries for Open Knowledge*, *22nd International Conference on Theory and Practice of Digital Libraries, TPDL 2018, Porto, Portugal, September 10–13, 2018, Proceedings*. Retrieved at https://arxiv.org/abs/1803.03198 (accessed on 15.8.2025).

Ignatowicz, J., Kutt, K., & Nalepa, G. J. (29. 5. 2025). Knowledge graphs for digitized manuscripts in Jagiellonian Digital Library application. *arXiv preprint*. Retrieved at https://doi.org/10.48550/arXiv.2506.03180 (accessed on 15.8.2025).

Summary

As the size and complexity of databases in memory institutions continue to grow, there is increasing interest in improving their searchability and reusability. From this perspective, the quality of metadata units is of great importance. High-quality metadata records should be precisely identified using concepts listed in validated authority files. This ensures consistency, accuracy, and machine-readability when describing people, places, organizations, or events. Each concept in an authority file should be represented by an authority record, that includes a unique, persistent identifier and a preferred name form.

When authority records contain additional information to further specify or contextualize the described concept, they go beyond simple identification. The collection of these enriched records forms what is called a namespace.

This paper explores the nature of namespaces, the logic behind their construction, the benefits of their use, and the potential for linking an institution's local namespace to global namespaces. It also presents practical challenges encountered in building a local geographical namespace at the National Archives of Hungary (NAH).

Typology: 1.04. professional article