UNLOCKING HISTORY: A REDESIGN OF THE SISTORY 5.0 PORTAL

Katja MEDEN,[1,3], Ana CVEK[1], Vid KLOPČIČ[2], Matevž PESEK[2], Mihael OJSTERŠEK[1], Mojca ŠORN[1], Andrej PANČUR[1]

[1]Institute of Contemporary History

[2]Faculty of Computer and Information Science, University of Ljubljana

[3]Department of Knowledge Technologies, Institute "Jožef Stefan"

The portal History of Slovenia - SIstory.si is an interdisciplinary collection of historical publications, data, collections and metadata that has been operating since 2008. The portal comprises a wide range of historical information, including publications, images and extensive databases as well as comprehensive metadata describing the objects. The recent redesign of the SIstory portal has focused on ongoing efforts to offer the data not only as a collection of historical publications, but also to enable greater transparency, interoperability and availability of research data to a wider audience. This paper presents the process of redesigning the portal in its technical and content enhancements, followed by an in-depth analysis of the content offered by the portal in its current form.

Keywords: SIstory, redesign, information systems, metadata, historiography

1 INTRODUCTION

The beginnings of the Research Infrastructure of Slovenian Historiography (RI INZ) date back to September 2008, when the online research and education portal of Slovenian historiography, *History of Slovenia - SIstory*, was launched. Its main content at the time consisted of a combination of historiographical research, historical sources and technical infrastructure outputs. Its main aim was to provide digitised and freely accessible research results and sources, with a focus on preserving older and not-so-easily accessible historical sources and thus preserving cultural and scientific heritage (Šorn et al., 2011). The operation and design of the SIstory portal were based on supporting the research process of the Institute's research community members. With the very successful initial response from the public and related institutions, SIstory

gradually moved beyond the boundaries of "written history" and developed interactive presentations of historical content, supported by the development of the then-new technologies (Šorn et al., 2011). While the content and initial design of the portal seemed to be synonymous with the concept of a "digital library", SIstory quickly became more than that. With the data preservation and integration of the research community at the forefront of SIstory's development, it became the main venue of a newly established Slovenian national node of the DARIAH research infrastructure family, DARIAH-SI, with the Institute (and more specifically RI INZ) as the national coordinating institution (Pančur & Šorn, 2019).

The redesign presented in this paper focuses on the technological advances that are now more accessible in order to improve the accessibility of the data. Furthermore, the data available on the portal has not yet been fully explored. Therefore, part of our work consists of a visual exploratory analysis of the collected data. Finally, we are focussing our efforts on the transparency and reuse of the data. In accordance with the European Commission's Data Act (*Data Act*, 2024), we have redesigned the portal to make all collected data fully accessible and to make the interaction more user-friendly.

The rest of the paper is structured as follows: Section 2 outlines the state of the portal before and after the redesign, focussing on the state of the portal, the reasons for the redesign and the components that were included in the redesign. Section 3 then provides some basic statistics about the content and presents the analysis of the portal content. Section 4 discusses the results, highlighting notable trends and their significance for the portal. Finally, Section 5 provides an overview of the paper and presents some options for future work.

2 HISTORY OF THE SISTORY

The SIstory portal can look back on a relatively long history of development. As already mentioned, the first steps were taken in 2008 with the initial release and since then several versions of the portal have been published as individual upgrades of the portal. In 2011, the first software and technological upgrade of the portal was carried out, which was a major step forward in Slovenian digital historiography. With this, SIstory not only established the latest standards and

enabled faster and more stable operation of the system, but also played an important role in establishing a national digital infrastructure for the humanities and arts (Cvek et al., 2022). The first upgrade consisted of several components (Rožman & Marolt, 2011):

- Content administration in SOLR[1] and upgrading folder structures and file names.
- Implementation of the Dublin Core metadata standard (the schema contained all 15 basic elements of DCMES,[2] a year later the original schema was upgraded with elements of the qualified DCMI Metadata Terms. (DCTERMS).[3]
- Creation of a unique and permanent URN—Uniform Resource Name.
- Introduction of the Sphinx metadata search engine.[4] Two search engines were implemented: basic and advanced.
- Upgrading of the portal administration.
- Design of the structure and access levels for users.

As the DCTERMS element set was no longer sufficient to adequately describe the different types of information sources, the portal was updated in 2013 to develop and integrate the SIstory metadata schema, a customised set of metadata elements that better matched the nature of the then-current content. The next major update of the portal took place in 2016. A mapping between the SIstory metadata schema and the DC was created to enable better interoperability of the data. The SIstory metadata schema was then extended with elements and structures from the HOPE application profile (a well-established profile in the GLAM community) (Lemmens et al., 2011) to develop the *SIstory application profile* (Pančur, 2013), which has remained the metadata set of choice for the SIstory portal over the years. The application profile and its implementation led to the structure, syntax and semantics of the metadata input tool.

Besides the metadata enrichment, new frameworks and graphical templates of the system, the graphical interface of both the administration and the user interface were also installed. In addition, the search engine (filtering and sorting of results; full-text search) was also taken into account in the updates (Cvek et

[1] Apache Solr
[2] Dublin Core Metadata Element Set
[3] DCMI Metadata Terms
[4] Sphinx search engine

al., 2022). Overall, since the portal's inception in 2008, a series of upgrades have been made, each improving the portal's functionalities and features. This brings us to today and to new steps in the portal's development – the decision to redesign the portal from the ground up.

3 SISTORY: THE REDESIGN

Immediately before the redesign, the basis of the portal consisted of various modules that were responsible for entering, processing and storing a collection of entities enriched with metadata. More specifically, the base consisted of:

- MySQL relational database
- the non-relational Elasticsearch database to search and index the content of collections and sources
- PHP 7
- File management module (extraction of raw text from PDF, HTML, XML, DOCX files, generation of previews)
- The metadata schema followed the previously mentioned SIstory application profile schema, developed for the purpose of flexible descriptions of various entities.

However, as the portal was updated several times, with each update based on the new versions of the same technology, this posed a problem as the code became too vast to manage efficiently. Furthermore, the concepts and various other solutions that were developed over the years were very ambitious and necessary for the time. Nevertheless, they did not prove to be as useful in the practical day-to-day operation of the portal as originally thought. This along with the outdated appearance of the user interface were decisive factors in our decision to start from scratch. When planning the redesign, we took into account the legacy issues and solutions from previous versions of the portal in order to improve the functionality of the system and provide a familiar user experience.

The redesign consisted of several sections, ranging from purely technical aspects (i.e. code base, integrating the OAI-PMH protocol) to simplifying the metadata schema, refining the user interface and restructuring the content classification.

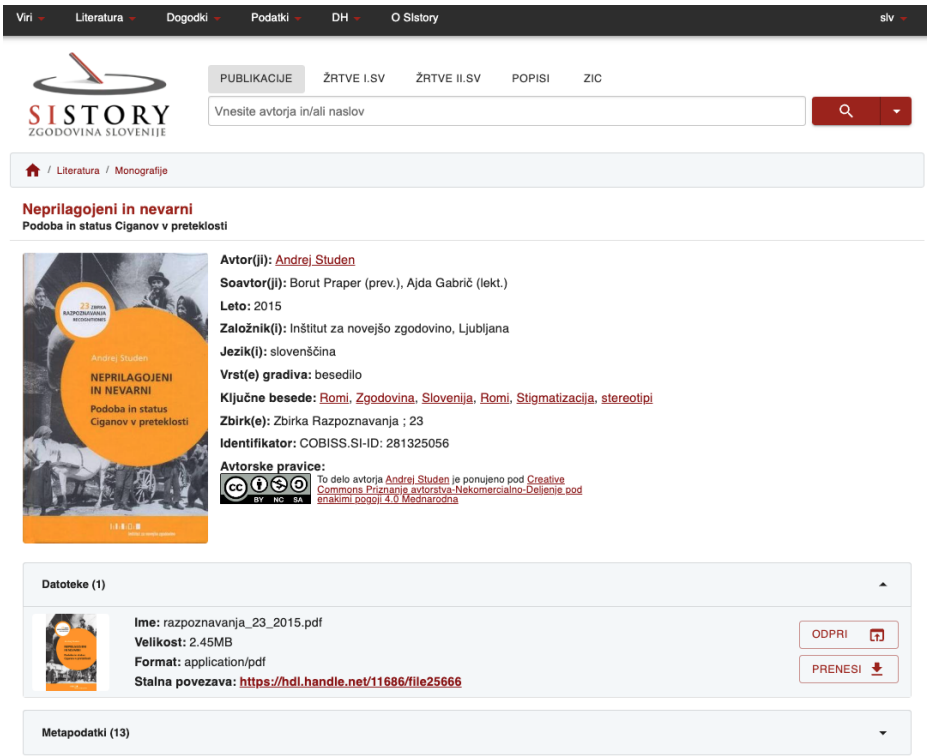

Figure 1: The new SIstory UI.

3.1 Technical design

In terms of the technical composition, the redesigned SIstory 5.0 portal is based on a robust technical framework, while the backend utilizes Django version 5.0.4, based on Python 3.11, for efficient data management and content delivery. The server infrastructure is supported by nginx version 1.25.

On the frontend, SIstory employs Next.js version 14.1.0 and React version 18.2.0 in combination with Node.js v18 for dynamic and interactive user interfaces. This modern frontend stack enables smooth navigation and responsive design across various devices and improves accessibility for users accessing historical content. Figure 1 shows an example of the redesigned UI.

The portal's database architecture is based on PostgreSQL version 16 and provides a robust foundation for storing and retrieving large volumes of historical

data with high speed and reliability. In addition, SIstory integrates Matomo version 5.0.3, an analysis function that provides administrators to gain valuable insights into user behaviour and interaction patterns, thus forming the basis for future developments and improvements.

For an efficient search functionality, SIstory incorporates Elasticsearch version 8.9.0 and Kibana version 8.9.0, so that users can quickly locate relevant historical documents and sources. The use of Elasticsearch ensures fast and accurate search results and improves the overall usability of the portal.

In addition, SIstory employs Handle[5] system to enable permanent identifiers that provide reliable and permanent access to specific historical documents and sources. This ensures that users can reliably reference and cite the materials, contributing to the scholarly integrity and reliability of the portal. Overall, SIstory's technical specifications underline the portal's commitment to providing a robust and user-friendly platform for accessing Slovenia's rich historical heritage.

3.2 Metadata design

The portal previously used the SIstory application profile (SIstory AP) as the basis for encoding the metadata, with the HOPE application profile serving as the basis. In practice, this posed a problem as SIstory AP contained several elements (and element groups) that were not used as frequently as originally assumed. This in turn led to a simplification of the profile. To this end, an analysis of the existing SIstory AP was carried out, with the main aim of identifying metadata elements that should be retained and addressing elements that present legacy issues.

The current state of the metadata application profile comprises 26 elements (reduced from the original 33 elements), with a focus on the DC and DCTERMS metadata elements and only a few additional elements from the previously mentioned HOPE AP. One of the main reasons for this shift in focus is to improve the interoperability of our metadata (i.e. DC is the base standard for the OAI-PMH metadata harvesting protocol).[6] A very limited number of elements of the

[5] Handle System identifier
[6] OAI-PMH protocol for SIstory 5.0 is at the time of writing still in the implementation phase.

namespace *sistory* have been retained,[7] mainly due to the remnants of the older publications described with these specific metadata elements. The overview of the major metadata groups is presented in Table 1.

Metadata	Unique Values	Nr. of Usages
No. of entries	57,263	
Creator	5,146	24,908
Subject	23,720	318,321
Publisher	1,036	53,699
Collection	426	1,861
Contributor	1,290	42,885
Type	12	58,868
Language	60	69,490

Table 1: Overview of the most important metadata groups, the number of unique instances and the total number of occurrences in the SIstory portal (at the time of creation).

In total, SIstory comprises over 57,000 unique entries, and over 5,000 unique authors/physical persons (under the category Creator),[8] while "*Subject*" contains the keywords that describe the publications. The secondary forms of authorship are described in the category "*Contributor*" (e.g. editor, translator...), while the type of publication on the basis of the controlled vocabulary (DCMI Type)[9] encompassing 12 categories. Finally, the portal includes publications in 60 different languages, which are presented in more detail in Section 4.3.

4 SISTORY UNVEILED: CONTENT ANALYSIS

In the efforts to present the redesign of the SIstory portal, it became clear that focusing on the mainly technical and aesthetic improvements would not fully capture the essence of the portal – its content, or rather, its historical sources. Therefore, we expanded the scope of the work to include a comprehensive

[7] For example, *SIstory Unstored* – an element/field for storing metadata that cannot be stored in any other metadata field due to its content.

[8] In the metadata mask there are two separate fields for a Creator, which according to the definition of Dublin Core can be either a physical person or a legal "organisation". Under the category "*Creator*" in the Table 1 only the occurrences for a physical person/author are counted.

[9] DCMI Type Vocabulary

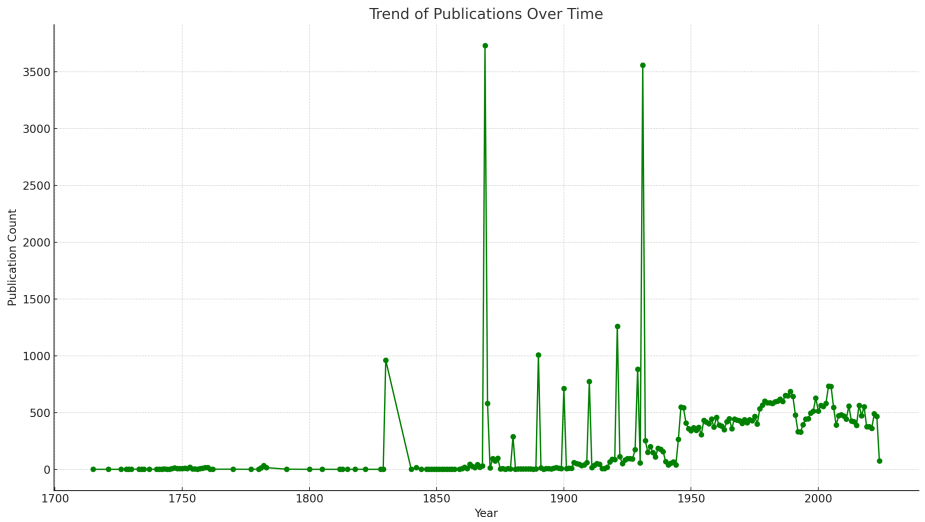

Figure 2: Trends of publications over time.

content analysis to gain a better understanding of the content available on the portal. In the analysis, we focused on different aspects of the portal's content, from the basic statistics of the main metadata groups (not the individual metadata elements) to analysing the keywords and language of the publications available on the portal. This in turn allows us to show not only the scope of content available, but also to highlight the different types and variations of this content.

In order to gain insight into the content available on the portal, we have focused on analysing the metadata of the publications. More precisely, we analysed the following specific metadata groups:

- Publication date (date when a specific publication was first published, i.e. in general, not on the portal itself),
- content menus (especially the first and second menu levels),
- language of the publication (according to ISO 639-2 standard),
- and the publication keywords (e.g. Jugoslavija, učbeniki...).

This in turn allowed us to capture some of the trends of publications for different time periods, more specifically, the distribution of publications over time, keyword analysis and language trends of publications.

4.1 Trends of publications

One of the trends analysed in this section is the distribution of publications over time according to their publication date, in order to check which years are best represented in terms of published content. The results are shown in Figure2.

One of the first trends to emerge is the distribution of publications within the period from around 1715 to the mid-20^{th} century, which showed several severe spikes in the number of publications. Conversely, the post-World War II period shows a much more even flow. The spikes in the timeline are most likely due to several large works or specific publications types (e.g. textbooks, population censuses) being published on the portal, while the steadiness of the publication flow in the post-war period to the present day seems to indicate a greater variance in the types of publications (e.g. literature, research and studies, monographs, etc.) and the absence of very large volumes of similar publications. The nature and content of these publication trends are examined in more detail in the keyword analysis (presented in Section 4.2) to provide additional insight and substantiate the reasons for the trends identified.

4.2 Keyword analysis

The keyword analysis of the portal content, in which the 10 most frequent/representative keywords used to describe the sources in the individual menus were examined for each of the 20 years (with the exception of the period 2010 - 2024). This was done for the first level menus: *Viri* (Sources), *Literatura* (Literature), *Dogodki* (Events), *Podatki* (Data) and *DH* (Digital Humanities) – the results are presented in the following sections.[10]

[10] Some of the keywords are very similar to one another, which is mostly due to slightly different versions of the notation format. For example, "*uradni list*" and "*uradni listi*" are singular and plural versions of the same keyword, but are counted separately.

4.2.1 SOURCES—TOP 10 KEYWORDS

Table 2 gives an overview of the 10 most frequent keywords found for each 20-year period for the menu "Sources". This category covers different types of resources, specifically archival, oral, and printed resources as well as digitized versions of physical objects. The latter are mostly images of physical objects, such as statues or death masks) and printed sources.
The keywords in the 18^{th} century period (or, more precisely, in the period from 1710 – 1790), the trend seems to consist of several repeating keywords, such as "*patenti*" (patents), "*odloki*" (decrees), "*norme*" (norms), "*Marija Terezija*", which refer directly to one of the more extensive collections of publications entitled "*Collection of various patents, decrees, ordinances, norms, instructions, etc., issued by Charles VI, Maria Theresa and Joseph II*",[11] which was acquired through our cooperation with the related institution, the Central Judicial Library. In addition to this collection, some of the keywords also refer to a large number of population censuses (as indicated by the keywords "*popisi prebivalstva*" (population censuses) and "*občina*" (municipality)) published for this particular period. This also applies to most of the 19^{th} century, where, in addition to the census, large amounts of theatre lists of various regional theatres (e.g. the Regional Theatre in Ljubljana), which in the past provided information on current plays or other related events. In addition to theatre lists and censuses, the calendars of the Society of St. Mohor (annual publication containing a calendar, religious prayers, illustrations, poetry, etc.)[12] were found in the keywords "*koledar*" or "*Družba sv. Mohorja*" (Society of St. Mohor), which were very popular at the time. Lastly, several keywords indicate a high number of minutes of the Carniolan Regional Assembly, as shown by the keywords "*Kranjska*", "*Carniola*", "*deželna avtonomija*" or "*provincial autonomy*".

At the beginning of the 20^{th} century, the most frequent publications uploaded to the portal are initially indicated with the already known keywords *popisi prebivalstva* and *občina*, which represent censuses, with several others, such as *uradni listi* (official gazettes), which refer to the official gazettes from different periods, republics and countries[13] (e.g. Slovenia, Yugoslavia, Serbia, Bosnia and Herzegovina, all of which are included in the keywords for this period) and

[11] Example of a Josef II directed patent
[12] Example of the St. Mohor calendar
[13] Official gazettes

Decade Range	Top Keywords
1710-1729	Karl VI., Patenti, odloki, predpisi, norme, navodila, okrožnice
1730-1749	Patenti, odloki, predpisi, norme, navodila, okrožnice, Marija Terezija, Karl VI., Karl VI., Corbinian Graf von Saurau, Marija Terezija, Anton Barbo Waxenstein, Marija Terezija, Anton Josef Auersperg, Marija Terezija, Anton Josepf Graf von Auersperg, Marija Terezija, Corbinian Graf von Saurau, Marija Terezija, Fridrich Wilhelm Graf von Haugwitz
1750-1769	Patenti, odloki, predpisi, norme, navodila, okrožnice, Marija Terezija, Karl VI., Marija Terezija, Marija Terezija, Anton Josepf Graf von Auersperg, Marija Terezija, Anton Joseph von Auersperg, Marija Terezija, Ludvik XVI., Marija Terezija I.
1790-1809	celjski grofje, drame, leposlovje, Celje, Ljubljana, hišne številke, rodbine, rokopisi, živinozdravniški recepti
1810-1829	Ljubljana, hišne številke, Ludvig van Beethoven, popis
1830-1849	Ljubljana, Slovenija, 1830-1857, popisi prebivalstva, programi, gledališča, 19. stoletje, gledališki listi, gledališče, Avstrija
1850-1869	Slovenija, 1869, popisi prebivalstva, Ljubljana, občina Dobrnič, občina Trebnje, občina Prečna, občina Mirna, občina Velika Loka, občina Črmošnjice
1870-1889	Slovenija, popisi prebivalstva, občina Vrhnika, 1870, 1880, Kranjska, deželna avtonomija, provincial autonomy, Carniola, koledar
1890-1909	Slovenija, popisi prebivalstva, občina Vrhnika, 1890, 1900, Družba sv. Mohorja, koledar, Avstro-Ogrska, popis prebivalstva, upravna razdelitev
1910-1929	Slovenija, popisi prebivalstva, Ljubljana, 1921, šolski listi, 1910, občina Vrhnika, 1929, Komunistična partija Jugoslavije, delavsko gibanje
1930-1949	Slovenija, Ljubljana, popisi prebivalstva, 1931, Jugoslavija, uradni listi, Srbija, BiH, Bosna in Hercegovina, uradni list
1950-1969	uradni listi, Jugoslavija, Ljubljana, BiH, Bosna in Hercegovina, Kosovo, Vojvodina, stenografski zapisniki, Socialistična republika Slovenija, družbeno samoupravljanje
1970-1989	Jugoslavija, uradni listi, stenografski zapisniki, predstavniška telesa, družbeno samoupravljanje, Socialistična republika Slovenija, Kosovo, BiH, Bosna in Hercegovina, Vojvodina
1990-2009	Slovenija, parlament, zakonodaja, državni zbor, Jugoslavija, uradni listi, skupščina, BiH, Bosna in Hercegovina, Vojvodina
2010-2024	popisi prebivalstva, Ljubljana, analiza, 1921, zgodovina, krajevna imena, 1900, krajevni leksikoni, toponimi, privilegiji

Table 2: Top 10 Keywords by Two-Decade Period for Viri (Sources).

stenografski zapisniki (stenographic records), which represent the minutes of various executive and legislative bodies, which were among the more prevalent publications and for which, the number of such publications only intensified within this period.

Finally, for the more recent period (2010 – 2024), the keywords refer mainly to studies carried out in connection with the censuses of Slovenia from 1830 – 1931, which are the result of cooperation with the Historical Archive of Ljubljana.

4.2.2 LITERATURE—TOP 10 KEYWORDS

Decade Range	Top Keywords
1810-1829	učbeniki, 19.st., abecedniki, slovenska književnost, slovensko-nemški abecednik, učbenik, učbeniki za osnovne šole, verouk
1830-1849	učbeniki, 19.st., izobraževanje, katekizem, katoliška vera, matematika, verouk
1850-1869	finance, Avstrijsko cesarstvo, učbeniki, slovnica, banke, valute, finančno vprašanje, slovenščina, valuta, nemščina
1870-1889	učbeniki, nemščina, matematika, politične stranke, organizacije in društva, čitanke, zgodovina, učbeniki za osnovne šole, berila, Kranjska, učbeniki za srednje šole
1890-1909	politične stranke, organizacije in društva, avstrijska doba, politični programi, Književna poročila, učbeniki, katoliški tabor, liberalni tabor, Narodopisne razprave in Mala izvestja, Mala izvestja, matematika
1910-1929	Slovstvo, politične stranke, organizacije in društva, politični programi, Izvestja, avstrijska doba, Razprave, učbeniki, liberalni tabor, katoliški tabor, zgodovina
1930-1949	Slovstvo, Razprave, Izvestja, zgodovina, učbeniki, Pregled, Zapiski, učbeniki za srednje šole, geografija, Jugoslavija
1950-1969	ocene in poročila, druga svetovna vojna, Slovenija, zgodovina, NOB, Ljubljana, zgodovinski pregledi, arheologija, Slovenci, Jugoslavija
1970-1989	ocene in poročila, druga svetovna vojna, Slovenija, arhivsko gradivo, arhivi, poročila, NOB, srednji vek, Jugoslavija, zgodovina
1990-2009	ocene in poročila, Slovenija, arhivi, druga svetovna vojna, zgodovina, arhivsko gradivo, Slovenci, arhivistika, biografije, Jugoslavija
2010-2024	ocene in poročila, Slovenija, zgodovina, Jugoslavija, druga svetovna vojna, socializem, Ljubljana, prva svetovna vojna, vojaška zgodovina, ocene

Table 3: Top 10 Keywords by Two-Decade Period for Literatura (Literature).

Similarly, Table 3 shows the 10 most frequent keywords for a single 20-year period for the Literature menu, which consists of publications such as research

monographs, (Slovenian) serial publications on historiography — together with the in-house produced scientific journal *Prispevki za novejšo zgodovino* (Contributions to Contemporary History)— school and university theses, and collections of digital monographs.

The 19^{th} century is predominantly dominated by the textbooks produced as part of the projects "*Šolski listi*" and "*Schools and Imperial, National, and Transnational Identifications: Habsburg Empire, Yugoslavia, and Slovenia*",[14] an extensive digitization project of textbooks primarily intended for schools on various school subjects identified in the table with the following keywords: *učbeniki* (textbooks), *abecedniki* (abecedarium), *matematika* (mathematics), *čitanke* and *berila* (reading material), etc. In the early to mid-20^{th} century, however, the topics are then expanded to include additional material on the topics of politics, political programmes and political parties, indicated by the keywords *politični programi* (political programmes), *katoliški tabor* (Catholic camp), *liberalni tabor* (Liberal camp). For the second half of the 20^{th} century, the themes shift to the Second World War, more precisely to the role of Yugoslavia (and Slovenia) in the Second World War (keywords). Directly related to this is also a considerable amount of literature referring to archival sources (keyword *arhivsko gradivo*)— mostly in connection with a specific journal, *The Gazette of the Archival Association and Archives of Slovenia*. Lastly, a very prominent keyword, *ocene in poročila* (reviews and reports), refers to a very specific form of contributions to various Slovenian (scientific) journals, in which the authors of the contributions give their reviews of various published works on the topics of the journal (in this case, mainly history).

4.2.3 EVENTS—TOP 10 KEYWORDS

While text documents are the predominant type of publication within the SIstory portal, RI INZ also offers in-house production and recording of various events and digitization of various exhibitions related to the field of historiography, the Institute or related institutions.

The first difference between the Tables 2 and 3 is the significantly shorter time period, which is not particularly surprising, however, the portal only exists from

[14] Schools and Imperial, National, and Transnational Identifications: Habsburg Empire, Yugoslavia, and Slovenia

Decade Range	Top Keywords
1990-2009	Kranjska, Ljubljana, Slovenija, dekleta, dokumentarni filmi, izobraževanje, univerze, zgodovina, študenti
2010-2024	Središče za javno zgodovino, Filozofska fakulteta, Oddelek za zgodovino, zgodovina, Slovenija, Jugoslavija, šolstvo, muzej, druga svetovna vojna, video

Table 4: Top 10 Keywords by Two-Decade Period for Dogodki (Events).

2008 onwards—this is also directly related to the relatively unrepresentative keywords within the period 1990 – 2009, as there are only a few individual publications related to this period (specifically, there are only 2 such publications). The topics of these publications are directly related to the topic of girls' education in Ljubljana[15] and Slovenian students abroad.[16] However, the number of publications increases in the period 2010– 2019. The most common keywords, such as *Filozofska fakulteta* (Faculty of Arts), *Oddelek za zgodovino* (Department of History), *zgodovina* (history) and Slovenija or Jugoslavija, refer to the institutions, organisations and general topics that organised the events (mostly recorded lectures).

4.2.4 DATA AND DH—TOP 10 KEYWORDS

In contrast to the keyword analysis of sources and literature, which covers several centuries, the two following publication types, Podatki (Data) and DH (Digital humanities data) are limited to the last decade (2010 – 2024). In both cases, the number of publications is relatively small, so these keywords are more representative of individual sources than of a large part of the portal.

Decade Range	Top Keywords
2010-2024	1910, Dravska banovina, Judje, Slovenije, krajevna imena, popisi prebivalstva

Table 5: Top 10 Keywords by Two-Decade Period for Podatki (Data).

[15] Šola naših babic: izobraževanje deklet v Ljubljani
[16] Študenti s Kranjske na avstrijskih in nemških univerzah 1365 – 1917

The Table 5 is mainly limited to research data in the field of historiography, more specifically, rich data on the old place names in Slovenia[17] and censuses of the Jewish population in Slovenia.[18]

While the categories examined are generally research data, the DH category relates primarily to the expanding vision of the RI INZ at the time – expanding into the field of digital humanities and providing data and tools to support research activity in these (related) fields. This later led to the development of a separate repository for digital humanities, the SI-DIH repository, another product of RI INZ and DARIAH-SI.

Decade Range	Top Keywords
2010-2024	nadgrajena resničnost, XML shema, SIstory augmented reality XML, metapodatki, DOCX, HTML publikacija, SIstory, SIstory nadgrajena resničnost XML shema, TEI, administracija

Table 6: Top 10 Keywords by Two-Decade Period for DH.

As these were the first steps of the infrastructure towards DH, there are only limited publications and tools available, but they incorporate the technologies of the time— this is also reflected in the keywords in Table 6, such as *nadgrajena resničnost* (augmented reality), *XML shema* (XML schema), *metapodatki* (metadata), *HTML*, *TEI*.

4.3 Language trends

In addition to the keyword analysis, we also took a look at the languages of the publications within the SIstory portal, as shown in Figure 3.

It is not surprising that the most frequent language of publications on the portal *History of Slovenia – SIstory.si* is Slovene, with a total of 46937 occurrences, most common in the period 1970 – 2010, especially between 1990 – 2010. The second most frequent language is Serbian (8072 occurrences), although an explicit distinction must be made here, as Serbian also belongs to two other language categories: Serbian (Cyrillic) for publications in Cyrillic script (912 in total) and the Bosnian/Croatian/Serbian category for publications where the language could not be explicitly identified (mostly publications referring to the *Official Gazettes* of Yugoslavia), which are the most frequent in publications.

[17] Place names in Slovenia 1.0
[18] List of Jews in Slovenia (Dravska banovina), 1937

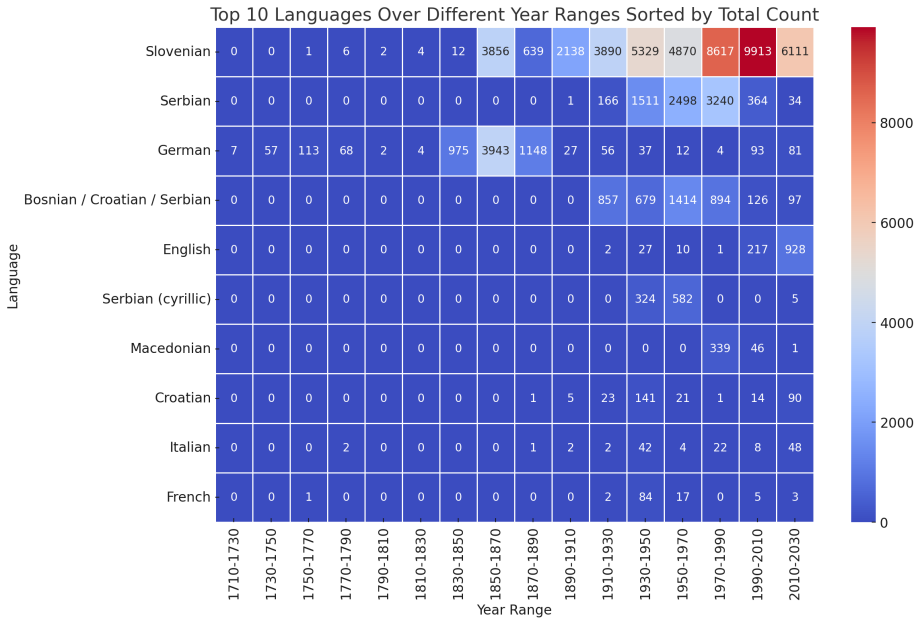

Figure 3: Language trends – distribution of publication languages over time.

Publications with all of the above categories for Serbian were published most frequently in the period from 1950 to 1990, again mainly in connection with the official g azettes. The last very frequent publication language is German, especially for the 19^{th} century, more precisely for the period 1830 – 1890, which includes older publications related to the history of Slovenia in the mentioned period, when Slovenia was a part of the Austrian Empire (between 1804 – 1867) and Austria-Hungary (1867 – 1918), where the publications were written in both German and Slovene. It is somewhat less prominent in the 18^{th} and the beginning of the 19^{th} century (1710 – 1830), but it is much more emphasised in relation to Slovene). While Table 4.3 highlights the 10 most common languages, the portal also includes publications (and languages) that are not so common (i.e. less than 5 times): Spanish - Castilian, Latin, Polish, Arabic, Albanian, Ukrainian, Esperanto. In addition, several publications are multilingual (42 publications), while some units of the portal have no language. This applies in particular to our digitised collection of death masks entitled "*The Casting of Death*".[19]

[19] The Casting of Death

5 CONCLUSIONS

In this paper, we have presented the technical, visual and content-related changes that have been implemented in the new SIstory 5.0 portal. The technical changes have been noted in the feedback from portal users as a better user experience due to its responsiveness and speed in terms of reduced page load.

The purpose of visualising the data in this paper served as a medium to show the current trends in data collection to bring the collected data closer to the general public. While the portal's content has recently been the focus of a study (Šorn & Cvek, 2023), the emphasis there was on the chronological additions to the SIstory portal over the course of its existence. Still, no metadata-based analysis has been conducted to help us understand the content coverage and themes of the portal. The redesign therefore presented us with the perfect opportunity to familiarise ourselves better with the content we have collected and worked on so far, thus creating a valuable foundation for the future.

The initial content study, based on trends of publications over time, provided an overview of the distribution of content and outlined the likely reasons for this. These were then further explored within the keyword analysis, which revealed to some extent that the type of publications within the peaks of the graph corresponded well with the hypothesised reason for such a distribution of content over time (i.e. large volumes of publications of the same type such as textbooks and censuses). This also applies to the language analysis, which mainly served to give us an overview of the variety of publication languages available on the portal. Given the historical context, it is not surprising that in the 18^{th} and 19^{th} century publications were predominantly in German, with some exceptions in Slovene, while in the 20^{th} century the language coverage started to expand to other South Slavic languages (again with Slovene as the dominant language).

Our future work will consist of an in-depth analysis of the internal technical process of data collection and a possible expansion of the scope of the metadata. The current state of the redesigned portal will serve as a basis for our future work, directly involving our community in the process, which has already been one of the cornerstones in the development of the portal. In the future, we want to involve our users directly in the development process from the very beginning

(and gather their feedback) regarding the planned future features. One of these possibilities is the integration of visualisations as tools that allow researchers and other users to more easily interact and work with the data within the portal. In addition, we will focus on existing collections to expand the data and provide an equivalent platform for accessible and reusable modern historical sources.

6 ACKNOWLEDGMENTS

The paper was written in the framework of the Slovenian Research and Innovation Agency-funded research infrastructure DARIAH-SI, the infrastructure programme I0-0013 “The Research Infrastructure of Slovenian Historiography infrastructure programme”, and research programme P6-0436 “Digital humanities: resources, tools and methods”.

REFERENCES

Cvek, A., Ojsteršek, M., & Šorn, M. (2022). *Izhodišča metapodatkovnih sistemov portala Zgodovina Slovenije – SIstory (2008–2016)*. Ljubljana: Inštitut za novejšo zgodovino. (ISBN 0 471 40300)

Data act. (2024). Retrieved 28. maj 2024, from https://digital-strategy.ec.europa.eu/en/policies/data-act

Lemmens, B., Janssens, J., Dyck, R. V., Bardi, A., Manghi, P., Beving., E., ... Straube, A. (2011). *Hope - the common HOPE metadata structure, including the harmonisation specifications (d2.2)* (Tech. Rep.). Project report, HOPE, Deliverable D2.2, 2011.

Pančur, A. (2013). *Metapodatki portala Zgodovina Slovenije - SIstory* (Tech. Rep.). Inštitut za novejšo zgodovino.

Pančur, A., & Šorn, M. (2019). Na začetku je bil SIstory: Raziskovalna infrastruktura slovenskega zgodovinopisja. In *Inštitut za novejšo zgodovino : 60 let mislimo preteklost* (pp. 47–58). Ljubljana, Slovenia. (https://hdl.handle.net/11686/46230)

Rožman, B., & Marolt, G. (2011). *Analiza podatkov in postavitev standardov in infrastrukture za nadgradnjo portala SIstory.si.*

Šorn, M., Pančur, A., & Sunčič, M. (2011). SIstory: arhivsko gradivo in e-humanistika. *Arhivi*, *34*(1), 145.

Šorn, M., & Cvek, A. (2023). *Vsebine in njihovo razporejanje na portalu Zgodovina Slovenije - SIstory (s poudarkom na publikacijah)*. Inštitut za novejšo zgodovino.

ODPIRANJE ZGODOVINE: PRENOVA PORTALA SISTORY 5.0

Portal Zgodovina Slovenije - SIstory.si predstavlja pomembno interdisciplinarno zbirko publikacij, podatkov, zbirk in metapodatkov, predvsem na področju zgodovinopisja. V zbirki portal zajema širok spekter zgodovinskih publikacij ter metapodatke, ki podatke opisujejo. Nedavna prenova portala SIstory je bila osredotočena na prizadevanja, da bi podatke ponudili ne le kot zbirko zgodovinskih publikacij, temveč tudi omogočili večjo preglednost, interoperabilnost in dostopnost raziskovalnih podatkov širšemu občinstvu, tako raziskovalcem kot laični javnosti. V tem prispevku je predstavljen proces prenove portala v njegovih tehničnih in vsebinskih izboljšavah, ki mu sledi poglobljena analiza vsebin, ki jih ponuja portal v sedanji obliki.

Keywords: SIstory, prenova, podatkovni sistemi, metapodatki, zgodovinopisje