

# Aggregation as a service

## Automatic topic detection and collaborative topic tagging in Archives Portal Europe's multilingual environment

Kerstin Arnold

**Biography.** Kerstin Arnold has been working in the archives domain for more than 15 years. During this time, she has gained expertise in a variety of areas, from academic editions via records management and transfer of born-digital materials to aggregation of cultural heritage metadata on national and international levels. Having been part of the projects establishing Archives Portal Europe, Kerstin is now the initiative's Manager. She holds a Master degree each in Communication Science and in Library and Information Management and also is a member of the Technical Subcommittee on Encoded Archival Standards. She shares her knowledge and experience with colleagues and with the next generation of archivists at conferences and by teaching at the University of Applied Sciences in Potsdam, Germany.

**Abstract.** Archives Portal Europe ([www.archivesportaleurope.net](http://www.archivesportaleurope.net)) is a comprehensive and open resource on archives from and about Europe, that currently holds archival descriptions from more than 30 countries and in more than 20 languages. Following traditional approaches of archival description, the portal allows users to access the documents via the contextual entities of the records creators and the holding repositories, next to a general keyword search. To evaluate options for subject- or topic-based access points, Archives Portal Europe is working on an automated cross-lingual topic detection tool that aims at enabling users to identify relevant documents related to a topic well beyond the narrowness of direct keyword matching. Synergising different approaches for concept-based and entity-based topics, the tool then also is meant to allow for active topic tagging in order to improve coverage of topic-based relations between the heterogeneous and multilingual documents present in Archives Portal Europe. Building on the current status quo in the portal, this paper presents the tool's set-up, initial results from the proof-of-concept phase, and next steps envisaged during alpha and beta development of the tool, which will be made available as Open Source to also be of benefit for other, similar initiatives in the cultural heritage sector.

### 1 Background

In the mid-2000s, the European Union saw its single largest enlargement, when ten countries joined the bloc on 1 May 2004, with two more being added on 1 January 2007. This change had its effects in all areas of life, including the archives domain. On 6 May 2003, the Council of the European Union had issued its resolution on archives in the Member States (OJ 2003/C113/2) (European Union 2003), resulting in the establishment of the European Archives Group (EAG). During the following two years, the EAG, in collaboration with the European Board of National Archivists (EBNA), worked on what should become the Report on Archives in the enlarged European Union (European Union 2005), adopted by the Council in its recommendation on priority actions to increase cooperation in the field of archives in Europe (OJ 2005/L312/55) from 14 November 2005. The report included recommendations for actions in different areas, with

“part two [...] deal[ing] with institutional, technical and professional aspects of access to archives. Particular emphasis is placed on [...] finding aids and archival description; access on line and new research tools; setting up an Internet Gateway/Portal to documents and archives in Europe and cooperation with European networking projects in this field [...].”  
(European Union 2005)

Thus, the idea for Archives Portal Europe ([www.archivesportaleurope.net](http://www.archivesportaleurope.net)) was born.

Funded by the European Commission in two rounds from 2009 to 2012 and again from 2012 to 2015, Archives Portal Europe is now managed and developed further by the Archives Portal Europe Foundation (APEF), who took over all responsibilities and rights from the APEx (Archives Portal Europe network of excellence, [www.apex-project.eu](http://www.apex-project.eu)) and APEnet (Archives Portal Europe network, [www.apenet.eu](http://www.apenet.eu)) projects in October 2015. The foundation has its physical headquarters in The Hague, Netherlands, but works with a remote and distributed core team of staff. It is supported financially by national archives, national archives administrations, and national archives aggregators from 21 countries in the role of foundation associates.

Archives Portal Europe is a comprehensive and open resource on archives from and about Europe, enabling new knowledge and new connections to be made. Its network represents a community of like-minded archives and cultural heritage professionals dedicated to the importance of sharing the continent’s shared history and heritage. At the time of writing, the portal aggregates descriptions of more than 600,000 archival collections, with the majority being described up to item level and including links to digital archival objects, where applicable. The content published on the portal comes from over 1,100 institutions located in more than 30 countries and is made available in more than 24 languages and currently five different alphabets.

## 2 Approaches to Archival Description

In the way in which Archives Portal Europe presents the content made available on its platform, it very much follows the traditional approaches of archival description: the agents involved with the archival documents throughout their lifecycle, the records creators, their activities and tasks, give a collection of archival documents their initial structure, classification, and grouping, which provides further insight in how documents relate to each other and hence is usually kept as is when transferring documents into the archives. The holding repository, on the other hand, will – while staying true to the original contextualisation – add further contextual information relevant to its function of preserving the archival records and making them accessible and available to the public.

The classic statement that “the archive arises as a consequence of the activities of the person who formed it [because] the documents can only be understood from the point of view of the task involved” (Muller et al. 1940/2003, xx-xxi) is furthermore extended by an hierarchical approach to archival description. Here, “the fonds forms the broadest level of description [and] the parts form subsequent levels, whose description is often only meaningful when seen in the context of the description of the whole of the fonds” (International Council on Archives 2000, 8). This adds to the complexity of navigating archival description that users are presented with, as the approach via a multilevel description also means that “information that is common to the component parts [is given at the highest appropriate level only and is] not

repeat[ed] at a lower level of description” (International Council on Archives 2000, 12).

Especially this last aspect of archival description leads to the question, how archives can embark on presenting the material they hold and the descriptions thereof in conjunction with and in relation to items from other cultural heritage institutions such as libraries and museums, where subject- and object-based approaches to description have traditionally been more common. And how archives can meet the expectations of users, who – in today’s digital world of information retrieval – have grown accustomed to searching by subject-based access points even more.

### 3 Traditional and New Access Points to Archival Materials

#### 3.1 Access Points in Archival Description Standards

While ISAD(G), the General International Standard Archival Description developed and maintained by the International Council on Archives “to be used in conjunction with existing national standards or as the basis for the development of national standards” (International Council on Archives 2000, 7), acknowledges the importance of access points for information retrieval in general, its main focus is on access points related to the agents that are named as records creators. With regard to other access points, the standard refers to national and language-specific developments as well as to more general conventions and frameworks that “are useful when developing and maintaining controlled vocabularies: ISO 5963 [...], ISO 2788 [...] and ISO 999 [...]” (International Council on Archives 2000, 9).<sup>1</sup>

Records in Contexts – A Conceptual Model for Archival Description (RiC-CM), the emerging new model for describing archives which is meant “to reconcile, integrate, and build on the four existing standards” (International Council on Archives 2012-2021)<sup>2</sup> extends the traditional approach by also including entities such as Event and Place, but defaults back to the general base entity Thing when it comes to connecting the archival documents with “all possible concepts, material things, or events within the realm of shared human experience and discourse [...] that are of primary interest to records managers and archivists, as well as other entities used in the description of the primary entities” (International Council on Archives 2021, 19). The accompanying Records in Contexts - Ontology (RiC-O), on the other hand, provides a whole range of additional classes for indexing archival description such as Type, Language, Physical Location, and Coordinates to only name a few.

#### 3.2 Access Points when Encoding Archival Descriptions

Following the existing international and/or national standards, conventions, and rules for archival description, most archival management systems will provide options to identify, name, and potentially describe agents who are of importance in the context

---

<sup>1</sup> For the ISO standards mentioned see: ISO 5963:1985, <https://www.iso.org/standard/12158.html>; ISO 2788:1986, revised by ISO 25964-1:2011, <https://www.iso.org/standard/53657.html>, and ISO 25964-2:2013, <https://www.iso.org/standard/53658.html>; ISO 999:1996, <https://www.iso.org/standard/5446.html>, all last accessed on 1 September 2021.

<sup>2</sup> Next to ISAD(G), the four standards mentioned here include: ISAAR(CPF) <https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd>; ISDF, <https://www.ica.org/en/isdf-international-standard-describing-functions>; ISDIAH, <https://www.ica.org/en/isdiah-international-standard-describing-institutions-archival-holdings>, all last accessed on 1 September 2021.

of archival documents; to certain extent, they will also include fields to capture information about other access points, such as subjects, on all levels of description, though these will often be bound to local or national vocabularies only.

The addition of such access points as part of the archival descriptions will also depend on other aspects such as whether or not it is part of the archival description tradition of an institution or country and the question of resources available to create detail-level descriptions to start with. Only in (versions of) archival management systems that have been developed more recently there will be a functionality that allows for the inclusion of references to international vocabularies such as the Library of Congress Subject Headings (LCSH), the Getty Art & Architecture Thesaurus (AAT), or the UNESCO Thesaurus.

### *3.3 Access Points when Aggregating Archival Descriptions*

The diversity of approaches also means that subject-based access points in archival descriptions present a specific challenge for aggregation initiatives, especially for those gathering materials from more than one country such as Archives Portal Europe.

Therefore, in addition to displaying existing subject headings as part of the archival descriptions and indexing these for the general keyword search in the portal, Archives Portal Europe has created a central, overarching topic-based approach that allows archival institutions to either make use of the appropriate encoding in their metadata or to add the relation between their archival documents and a specific subject as part of the central data processing. By this, users are offered access to the archival material via a list of predefined subject terms that connect documents in a variety of languages based on the mechanisms set out in the back-end of Archives Portal Europe.

However, the current process is not very flexible with regards to extending the list of subject terms that are available, and it furthermore relies on a manual intervention by the contributing institutions themselves during data processing, which leads to a rather inconsistent representation and coverage of these central topics across all countries and thereby languages connected to Archives Portal Europe at the moment.

## **4 Automated Topic Detection in a Multilingual Environment**

Given the status quo as described in the previous section, Archives Portal Europe, in collaboration with an external developer and data scientist and supported by King's College London, has initiated a Research & Development project back in 2020 to evaluate the possibilities of applying automatic topic detection to its multilingual environment. Following the promising results of the initial proof-of-concept phase (see chapter 4.5), this work is currently being extended in the context of Archives Portal Europe's contribution to the Europeana Digital Service Infrastructure.

### *4.1 Objectives*

The aim is to exploit methods of Natural Language Processing (NLP) in a supervised approach in order to train a tool that will help users as well as contributors to Archives Portal Europe in identifying materials related to a topic of their interest. In the short-term, this will be based on the topics that already exist in the central system of Archives Portal Europe, but it is the intention to also enable the work on

and creation of new topics in the medium-term with the help of automated topic detection.

In the long-term, it is envisaged to add a second step following on from the identification of relevant documents: the possibility for contributors to Archives Portal Europe as well as its users to flag up these documents for active tagging with an appropriate subject heading, either in the central system or – ideally – at the source of the data, using international Linked Open Data vocabularies as a basis.

While developed in the context of and with the data aggregated by Archives Portal Europe, the approaches followed and the functionalities provided by the final tool, which will be made available as Open Source, are meant to be applicable to any heterogeneous and multilingual dataset from the cultural heritage and related domains.

#### *4.2 Research Background*

The project builds on work conducted in the Digital Humanities during the past two decades with a focus on adopting NLP methods for identifying topics in a supervised approach. This technique has been found suitable based on the circumstance that Archives Portal Europe (1) already provides a predefined list of topics, (2) works with a set of materials that have been manually annotated with the relevant topics, and (3) comes with a relatively large amount of such pre-annotated materials, approximately 2 million documents in total, to train the tool on documents in different languages.

Especially the last point qualifies the Archives Portal Europe case for a supervised approach, for which having a large enough dataset for training purposes often is the biggest obstacle, while the method generally offers a reliable performance for topic detection tasks (see for instance the experiments conducted by Merz et al. 2016 and by Glavaš et al. 2017 on the Manifesto Corpus). The alternative, i.e. an unsupervised approach, has been ruled out on the basis of the first and second points mentioned above: in a case like the one of Archives Portal Europe, in which the user already knows the topics contained in the collection, it can prove difficult to employ an unsupervised approach (Owens 2012), especially when applying Latent Dirichlet Allocation topic models, the results of which are often extremely hard to interpret (Chang et al. 2009) and are not always straightforward to align with our common notion of topics.

#### *4.3 Cross-lingual Topic Classification for Concepts*

Special emphasis has been given to supervised set-ups using a Support Vector Machine (SVM) (Joachims 2002) to train an algorithm, where each document is represented as a single feature-vector capturing the “meaning” of its content. In the case of Archives Portal Europe, we consider the descriptive units<sup>3</sup>, i.e. the constituent components of an archival finding aid, as such “documents”, representing each of them as the averaged vector of all its words, and thereby obtaining a single “document embedding” for each description. In order to address the unbalanced

---

<sup>3</sup> The tool uses the Solr results in JSON format for each of these “documents”, where some major parts of the archival description are captured in singular fields (e.g., the title of the unit itself or of the upper hierarchical levels that this unit is a part of). However, other parts of the archival descriptions are only included in a placeholder field of the Solr index, capturing all additional metadata that might be part of the original EAD-XML file. This is currently not part of the “document” as used by the tool.

representation of languages in the dataset (see section 3.3), we apply Fast-Text word-embeddings and align these in a common cross-lingual “semantic” space by the project MUSE (Conneau et al. 2017-2018) to better represent all languages present in our dataset<sup>4</sup>. This approach has achieved really high performance, identifying the correct topical label for the materials in over 90% of the cases.<sup>5</sup> We additionally have ensured that the classifier was correctly distinguishing between topics, and not languages, by conducting an in-depth error analysis.

While the proof-of-concept tool only allowed for single term keywords, this has been extended during the alpha phase to enable searches with the Boolean operators AND, OR, and AND NOT, as well as the use of wildcards like the asterisk replacing one or more letters of a search term. In such search scenarios, the tool will also show which terms have been included based on regular expressions running in the background while the search is conducted. It should be noted that at the time of writing this article, this functionality had only been added as a new extension to the tool and hence will require some more detailed testing to confirm its impact.

#### *4.4 Extension to Entity-based Topics*

In addition to searching for concepts, the tool also offers the option to search for entities across languages. Instead of relying on cross-lingual embeddings, the retrieval function first maps the entity inserted by the user as a query to its equivalent in Wikidata (when present). Next, it retrieves all name variations in the other languages under study,<sup>6</sup> and finally searches for their occurrence in the corpus. In the most recent alpha phase, this has been extended to also connect to the Virtual International Authority File, VIAF, and to include additional name variations from there. While this function is an early prototype and does not fully rely upon entity disambiguation approaches yet, we consider it useful as an additional way of exploring the collection.

#### *4.5 Initial Results from the Proof-of-concept Phase*

Tests with the proof-of-concept version of the tool have used a set of 457,538 documents already tagged with 1 of 9 topics<sup>7</sup> selected based on the following criteria:

- Balancing topics that cover a language broadly enough to learn from with topics that include documents in more than one language in order to address the multilingual character of Archives Portal Europe;
- Having topics of varied size and scope;

---

<sup>4</sup> The dataset used in the proof-of-concept phase included documents in Finnish, French, German, Latvian, and Polish. Furthermore, we added English and Italian as supported languages for user queries. In the alpha phase, April to August 2021, this was extended further to also include Hebrew, Latvian, Russian, Spanish, Swedish.

<sup>5</sup> We obtained over 0.9 of both micro and macro F1-score, which is the harmonic mean of precision and recall. To know more see the documentation of the metrics on Scikit-learn, the library we adopted: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision\\_recall\\_fscore\\_support.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html).

<sup>6</sup> We pre-process name variations leaving life dates aside for persons or other characteristics sometimes included in brackets.

<sup>7</sup> For the alpha phase the data set was extended by four additional topics, reaching a total of about 675,000 documents being included.

- Including topics that are entity-based as well as topics that are concept-based to address the two main approaches in archival research: persons/places on the one hand, subjects/themes on the other.

With regard to both, the concept search and the entity search, the evaluation of the tool's predominant function at this initial stage, the discovery of documents relevant to a topic, has given promising results. The tool allows the user to identify relevant documents related to a topic well beyond the narrowness of direct keyword matching and it has shown good results as well from not very largely annotated topics, which opens the door for smaller scale projects in future.

Nevertheless, it should be noted that, with the test data set only representing a rather small sample of the repository of Archives Portal Europe (25% of all documents tagged with a topic, and 0.16% overall), our experiments are only a first attempt towards a very challenging goal, and we plan to work with an interaction of supervised and unsupervised methods in future experiments, in order to tackle these challenges in a more comprehensive way.

## 5 Next Steps

Apart from the confirmation of the proof-of-concept, the initial results also have allowed us to elaborate on areas of further investigation and future development. Based on this, we have enlarged the sample data both in terms of topics under consideration and of available languages for the alpha phase, have enabled Boolean operators and wildcards in the search functionalities of the tool, and have started with the integration of other vocabularies and ontologies next to Wikidata along with an initial option for entity disambiguation.

The tool has been redesigned as a web application<sup>8</sup> including the display of other data from the documents, e.g. dates and the country where the contributing institution is located, that might be useful in determining whether a search result is relevant to a specific topic of interest or not. The tool now also allows for an export of results in CSV format for further analysis offline.

One of the next steps will include more detailed testing of these latest extensions and new functionalities in the context of workshops to be held with members of the archives community as well as with researchers in late 2021 and in early 2022. These workshops might concentrate on a specific topic or a certain language or language family in order to extend their representation in the Archives Portal Europe's data set, or they might test the tools functionality more generally without predefining a topic or language context.

In terms of developments, the focus will be on the inclusion of further vocabularies and ontologies such as the LCSH, the Getty AAT, or GeoNames, and the extension to the full data set of Archives Portal Europe. Furthermore, the beta phase will look into making use of the tool's results from the aggregating perspective of Archives Portal Europe itself, i.e. with regard to enabling Linked (Open) Data connections based on the entities identified via the tool and with regard to making such enriched metadata available in some way to the portal's users and contributors, and from the perspective of a contributor to Archives Portal Europe, i.e. with regard to

---

<sup>8</sup> The public version of the tool is available at <http://topicdetection.archivesportaleurope.net/>. This will not always show the latest status of the development, but represents the most current stable version. All developments can be followed on GitHub at <https://github.com/ArchivesPortalEuropeFoundation/Topic-Detection/>.

transforming the results brought back by the tool into actual topic taggings in order to increase the representation of subject- or topic-based relations in the source data.

## References

1. Chang, Jonathan, and Gerrish, Sean, and Wang, Chong, and Boyd-Graber, Jordan L., and Blei, David M.. 2009. "Reading Tea Leaves – How Humans Interpret Topic Models." *Advances in Neural Information Processing Systems 22 (NIPS 2009)*. <http://papers.nips.cc/paper/3700-Reading-tea-leaves-how-humans-interpret-topic-models.pdf> (last accessed on 1 September 2021).
2. Conneau, Alexis, and Lample, Guillaume, and Ranzato, Marc'Aurelio, and Denoyer, Ludovic, and Jégou, Hervé. 2017-2018. *Word Translation Without Parallel Data*. <http://arxiv.org/abs/1710.04087> (last accessed on 1 September 2021).
3. Europeana. 2018-2021. *Europeana Digital Service Infrastructure*. <https://pro.europeana.eu/project/europeana-dsi-4> (last accessed on 1 September 2021).
4. European Union. 2003. *Council Resolution of 6 May 2003 on archives in the Member States*. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2003:113:0002:0002:EN:PDF> (last accessed on 1 September 2021).
5. European Union. 2005. *Report on Archives in the enlarged European Union*. <https://ec.europa.eu/transparency/regdoc/rep/1/2005/EN/1-2005-52-EN-F1-2.Pdf> (last accessed on 1 September 2021).
6. European Union. 2005. *Council recommendation on priority actions to increase cooperation in the field of archives in Europe*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005H0835&from=EN> (last accessed on 1 September 2021).
7. *GeoNames*. <https://www.geonames.org/>. (last accessed on 1 September 2021).
8. *Getty Art & Architecture Thesaurus*. <https://www.getty.edu/research/tools/vocabularies/aat/index.html> (last accessed on 1 September 2021).
9. Glavaš, Goran, and Nanni, Federico, and Ponzetto, Simone Paolo. 2017. "Cross-Lingual Classification of Topics in Political Texts." *Proceedings of the Second Workshop on NLP and Computational Social Science*: 42–46. Vancouver: Association for Computational Linguistics.
10. International Council on Archives. 2000. *ISAD(G) – General International Standard Archival Description. 2nd edn*. Ottawa. [https://www.ica.org/sites/default/files/CBPS\\_2000\\_Guidelines\\_ISAD%28G%29\\_Second-edition\\_EN.pdf](https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf) (last accessed on 1 September 2021).
11. International Council on Archives. 2004. *International Standard Archival Authority Records – Corporate Bodies, Persons, and Families (ISAAR(CPF)). 2nd edn*. Paris. [https://www.ica.org/sites/default/files/CBPS\\_Guidelines\\_ISAAR\\_Second-edition\\_EN.pdf](https://www.ica.org/sites/default/files/CBPS_Guidelines_ISAAR_Second-edition_EN.pdf) (last accessed on 1 September 2021).
12. International Council on Archives. 2007. *International Standard Description of Functions (ISDF). 1st edn*. Paris. [https://www.ica.org/sites/default/files/CBPS\\_2007\\_Guidelines\\_ISDF\\_First-edition\\_EN.pdf](https://www.ica.org/sites/default/files/CBPS_2007_Guidelines_ISDF_First-edition_EN.pdf) (last accessed on 1 September 2021).



13. International Council on Archives. 2008. *International Standard Description of Institutions with Archival Holdings (ISDIAH)*. 1st edn. Paris.  
[https://www.ica.org/sites/default/files/CBPS\\_2008\\_Guidelines\\_ISDIAH\\_First-edition\\_EN.pdf](https://www.ica.org/sites/default/files/CBPS_2008_Guidelines_ISDIAH_First-edition_EN.pdf) (last accessed on 1 September 2021).
14. International Council on Archives. 2012-2021. *Records in Contexts – Conceptual Model (RiC-CM) Homepage*. <https://www.ica.org/en/egad-ric-conceptual-model> (last accessed on 1 September 2021).
15. International Council on Archives. 2017-2021. *Records in Contexts – Ontology (RiC-O)*. [https://www.ica.org/standards/RiC/RiC-O\\_v0-2.html](https://www.ica.org/standards/RiC/RiC-O_v0-2.html) (last accessed on 1 September 2021).
16. International Council on Archives. 2021. *Records in Contexts – Conceptual Model v0.2*. [https://www.ica.org/sites/default/files/ric-cm-02\\_july2021\\_0.pdf](https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf) (last accessed on 1 September 2021).
17. ISO 999:1996, *Information and documentation – Guidelines for the content, organization and presentation of indexes*. <https://www.iso.org/standard/5446.html> (last accessed on 1 September 2021).
18. ISO 5963:1985. *Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms*. <https://www.iso.org/standard/12158.html> (last accessed on 1 September 2021).
19. ISO 25964-1:2011, *Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval*. <https://www.iso.org/standard/53657.html> (last accessed on 1 September 2021).
20. ISO 25964-2:2013, *Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies*. <https://www.iso.org/standard/53658.html> (last accessed on 1 September 2021).
21. Joachims, Thorsten. 2002. *Learning to Classify Text Using Support Vector Machines*. Boston: Springer.
22. *Library of Congress Subject Headings*. <https://id.loc.gov/authorities/subjects.html> (last accessed on 1 September 2021).
23. Merz, Nicolas, and Regel, Sven, and Lewandowski, Jirka. 2016. “The Manifesto Corpus – A new resource for research on political parties and quantitative text analysis.” *SAGE Journals, Research & Politics*.  
<https://journals.sagepub.com/doi/10.1177/2053168016643346> (last accessed on 1 September 2021).
24. Muller, Samuel, et al. 1940. *Manual for the arrangement and description of archives*. 2nd edn. Translated by Arthur H. Leavitt. New York: The H.W. Wilson Company. - Reissued 2003. Chicago: Society of American Archivists.
25. Owens, Trevor. 2012. *Discovery and Justification are Different – Notes on Science-ing the Humanitie*.  
<http://www.trevorowens.org/2012/11/discovery-and-justification-are-different-notes-on-sciencing-the-humanities/> (last accessed on 1 September 2021).
26. *UNESCO Thesaurus*. <http://vocabularies.unesco.org/browser/thesaurus/en/> (last accessed on 1 September 2021).
27. *Virtual International Authority File*. <https://viaf.org/> (last accessed on 1 September 2021).