

The Benefits of Linking Metadata for Internal and External users of an Audiovisual Archive

Victor de Boer^{1,2}, Tim de Bruyn^{1,2}, John Brooks^{1,2} and Jesse de Vos²

¹ Vrije Universiteit Amsterdam, the Netherlands

² Netherlands Institute for Sound and Vision, Hilversum, the Netherlands
v.de.boer@vu.nl

Abstract. Like other heritage institutions, audiovisual archives adopt structured vocabularies for their metadata management. With Semantic Web and Linked Data now becoming more and more stable and commonplace technologies, organizations are looking now at linking these vocabularies to external sources, for example those of Wikidata, DBPedia or GeoNames. However, the benefits of such endeavors to the organizations are generally underexplored. In this paper, we present an in-depth case study into the benefits of linking the “Common Thesaurus for Audiovisual Archives” (or GTAA) and the general-purpose dataset Wikidata. We do this by identifying various use cases for user groups that are both internal as well as external to the organization. We describe the use cases and various proofs-of-concept prototypes that address these use cases.

Keywords: audiovisual metadata, Wikidata, case study

1 Introduction

The Netherlands Institute for Sound and Vision (NISV)¹ is an audio-visual archive located in Hilversum, the Netherlands. This cultural historic institute manages over 70% of all material relating to Dutch AV heritage including Dutch radio programs, television programs, documentaries and music. Together with several other Dutch organisations who manage Dutch audio-visual heritage, NISV have developed the GTAA, the Common Thesaurus for Audio-visual Archives². This thesaurus provides the terms used in part of the metadata for the audio-visual material. The GTAA includes person names, objects, subjects, genres, geographical data, concepts, company names and so on. It also contains synonyms, quasi-synonyms, homonyms and different spellings for these terms.

With the help of linked data, connections can be made between the own collection and other data sources [1]. Cultural heritage institutes like NISV are looking for new ways to manage their cultural heritage collections and are exploring the possibilities of using Linked Data to add value to their collections [2]. Cross-linking between organization- or domain-specific vocabularies and more general purpose linked data

¹ <http://beeldengeluid.nl>

² <http://gtaa.beeldengeluid.nl>

sets allows institutions to enhance their collection with a vast amount of additional data. This promise of Linked Data, coupled with the maturity and uptake of the technology has convinced many organizations of the potential added value of linked data. However, in many cases, the actual benefits to end-users remain unclear and under-explored. In this paper, we investigate the added value using four very specific use cases. We specifically do this for internal users and external users of the metadata. Internal users include employees and researchers inside the NISV organization, while external parties are heritage professionals, media professionals, teachers, students and researchers, each with their own information needs.

In each case, we start with a use case elicitation process based on interviews with the stakeholders. This process then results in a specified use case, where links between GTAA and the general-purpose knowledge base Wikidata [3] are exploited for a specific (information retrieval) task. We then develop a proof-of-concept tool or mockup that matches the use case and finally perform a limited evaluation to show the added value and receive additional feedback on the case. This work is limited to the part of GTAA describing persons, but similar cases can be constructed for other parts of GTAA.

In Section 3, we provide a brief analysis of GTAA, Wikidata and existing links for the person concept scheme. Then, in sections 4 and 5 we describe use cases for internal and external users respectively. In each section, we describe the elicitation method, the use cases themselves as well as the proof-of-concept end-user applications. We close this paper with a short discussion on generalized findings. In the next section, we first discuss related work.

2 Related Work

Vrandečić & Krötzsch state that Wikidata's goal is to allow its data to be used both internally and in external applications, such as the external thesauri [3]. Erxleben et al. have conducted research on Wikidata and its possibilities for being connected to the linked data [4]. They describe how Wikidata is linked to multiple external datasets. These existing links range from the ISSN dataset, which identifies all journals and magazines etcetera, to more highly specialized databases, such as the database of North Atlantic hurricanes. Färber et al. looked specifically into the quality of the most noteworthy large knowledge databases including Wikidata [5]. This extensive research show that multiple quality aspects of the data contained within Wikidata were the highest of any of the noteworthy knowledge graphs. This research provides motivation for our own research to investigate links to Wikidata.

Thornton et al. also conducted research on Wikidata's possible role to serve as repository for international organizations concerned with digital preservation [6]. They concluded that Wikidata had the advantage of being structured, query-able and computable. Another advantage Wikidata has is that it's multilingual. Since Wikidata also had a Dutch version it supports alignments to the GTAA better.

Debevere et al. conducted research into linked data to improve the metadata of thesauri. This improved metadata allows better retrieval of information on search que-

ries. To this end, they developed an alignment algorithm that automatically linked data categories of DBpedia to another media thesaurus used to annotate archived media items at the Flemish public service broadcaster (VRT). It returned acceptable results, specifically for the Person category [7].

Tordai et al. provide a systematic approach to build a large semantic culture web [8]. They did so by making clear to heritage institutions what they need to do to make their collections fit for becoming a part of this semantic culture web. They advise to use the paradigm of open access, open data and open standards. They conclude that collection owners should be provided with the necessary support facilities.

A similar case study approach as ours was described in [9]. This paper describes the linking of a thesaurus for economics to DBpedia and other large thesauri using SKOS/RDF methods. In the end the linking to DBpedia returned a high number of unsuccessful matches. These unsuccessful matches are contributed to simple derivations in the compared strings and the fact that a significant number of economic concepts did not exist in DBpedia yet.

3 Analysis of Linked GTAA and Wikidata

GTAA. The Common Thesaurus for Audiovisual Archives (GTAA) is used by NISV to annotate the different collections. It follows the ISO-2788 standard for thesaurus structures and consists of several *concept schemes* the media objects: subjects, people, entities, locations, etc. The GTAA, available as SKOS, contains approximately 180,000 terms and is actively maintained, being updated as new concepts emerge on television. It includes about 90,000 scope notes.

Wikidata. Wikidata is continuous crowd-sourced concept where Wikipedia's data is cleaned and integrated to make it machine-readable as well as human-readable [3]. It is available in RDF and can be queried using the SPARQL query language. Each term or object within Wikidata has a unique identifier (URI), and several other identifiers from different sources. Each term has its own label, aliases and description, including property-value pairs and qualifiers, which includes contextual information.

Alignment between GTAA and Wikidata Persons. A partial alignment between Wikidata and GTAA was established by the Wikidata community using the "Mix 'n' Match" tool³. This tool was used to automatically match person entities based on first and last names. In total, 10,350 GTAA person concepts are linked to Wikidata entries. Even though this is only a small percentage of the total number of person names in GTAA (8.4%), this includes the most well-known people. An example of a linked person concept is found at <https://www.wikidata.org/wiki/Q37079>, which shows the Wikidata information for actor Tom Cruise, including the link to the matching GTAA concept <http://data.beeldengeluid.nl/gtaa/86659>. If we can identify enough added benefits, the scope of the alignment can be increased through further alignment initia-

³ <https://tools.wmflabs.org/mix-n-match/#/catalog/34>

tives. To identify the added value of these Wikidata links, we determine statistics on a number of biographical properties found in Wikidata. Table 1 shows the coverage for each of these properties. This shows that basic information is generally available, although in some cases, potentially interesting properties are missing. In accordance with [10], we found that generally, popular entries have a high completeness whereas less popular entries have a low completeness.

Table 1. Coverage for selected biographical properties for linked GTAA concepts in Wikidata

Property	Occurrences	% of total
Name	10,350	100.00%
Gender	10,294	99.46%
Date of birth	10,040	97.00%
Occupation	9,988	96.50%
Country of citizenship	9,976	96.39%
Place of birth	7,390	71.40%
Date of death	3,644	35.21%
Place of death	2,776	26.82%
Birth name	839	8.11%
Pseudonym	124	1.20%

4 Use Cases for Internal Users

4.1 Elicitation

A series of interviews with stakeholders of different departments of NISV was conducted. These departments deal with media 1) intake, 2) information management and 3) research and development respectively. The goal of these interviews was to result in use cases for the linked data. The interviewee is shown the extra data contained within Wikidata. The interviewee is also shown the results of the earlier analysis. After, the interviewee is asked if he/she based on this data sees any concrete use cases for when more data gets linked. This resulted in three use cases which we further explore in subsequent sections. In Section 4.5, we describe the evaluation of the three use cases.

4.2 UC-I-1: Receiving an alert when the copyright on a person's work expires.

This use case was identified by interviewees of all departments and represented a suggested use of the links that was earlier discussed within the organization. This use case is that of a copyright expiration alert on the work of persons in the GTAA came forward. In Dutch copyright law, works enter the public domain 70 years after a maker's death. On the basis of alignments between the GTAA and Wikidata, using the date of death can be retrieved for specific persons. The use case involves for information management and other stakeholders to get an alert on first day of a year, 70 years after a person's death. The user can then review the information and determine whether works by that person can be made openly accessible to the public [11].

Table 1 shows that 35.21% of the investigated subset have indeed a date of death available in Wikidata. Based on these statistics, we can determine that we should at maximum expect a return of 3,644 people for the alert. This number was further reduced by explicitly focusing on people that worked as producers/creators rather than other people in the thesaurus (for example politicians or sports persons who *appear* in programs). To this end, we filter on occupations. In total we found that there were 798 different occupations in Wikidata and selected manually those that are related to the fields of television, movies, radio, theater and music. This results in a reduced set of 1,626 persons.

To receive alerts, a simple proof-of-concept application was developed which generates calendar events for each of the potential out-of-copyright dates. Stakeholders can consult this calendar or subscribe to it. Specifically, as NISV uses Google for their email, calendar and daily activities and therefore a Google Calendar was created. Figure 1 shows an example of a copyright expiration notice for January 2030



Fig. 1. Example of a copyright expiration notice for January 2030 in Dutch.

4.3 UC-I-2: Provide more information on a person appearing in online story

The second use case that came forward during the interviews was the ability to provide extra information on persons mentioned in the new story platform currently under construction: <https://www.beeldengeluid.nl/verhalen>. This platform is a news platform maintained by the institute and contains multiple stories (blog posts). In the current version, the author of the story is already displayed in a sidebar. Using the data in Wikidata it is possible to present quick background information on the persons mentioned in the story.

Figure 2 shows a screenshot of the mockup we created, where mentioned persons are displayed in the story are now added to that sidebar. When someone clicks on one of the persons on the bottom of the story he/she will be redirected to a separate webpage generated from Wikidata containing extra information on said person as well as other stories the person is mentioned in.

DE KONING GAAT MET ZIJN TIJD MEE

Met Willem-Alexander als koning lijkt er een nieuw tijdperk aangebroken voor de relatie tussen het Koningshuis en de media. De koning schuwt grote interviews niet en op Dumpert is zelfs een 'dab' van hem terug te vinden. Zijn moeder en oma hadden er destijds wat meer moeite mee. Jeroen Snel, koningshuisverslaggever en presentator Blauw Bloed, legt uit hoe de relatie tussen het Koningshuis en de media in elkaar zit.



Verhaal

 **Jeroen Snel**
Verslaggever Koninklijk Huis

Personen

 **Willem-Alexander der Nederlanden**
Koning der Nederlanden

 **Beatrix der Nederlanden**
Koningin der Nederlanden

 **Juliana der Nederlanden**
Koningin der Nederlanden

BELEN

Fig. 2. Screenshot of the mockup made for use case 2. This shows persons mentioned in an example article on the NISV story platform.

In a finalized version of the tool, story authors can be suggested persons to be displayed based on Entity resolution results and select relevant person entities as a fully automatic tool could potentially display information on many persons that are not relevant to the story.

4.4 UC-I-3: Using Wikidata for story recommendation.

This third use case also involves the story platform of the previous case. When persons are mentioned in the story this can be used to generate “Stories you might also like” links. Such recommendation can be done on the basis of structural properties retrieved from Wikidata. An example could be about a story about US news anchor Walter Cronkite⁴. Where Wikidata includes many properties that can be used for such recommendation (for similar linked data based recommendation strategies see [12]).

At the same time, many properties are less relevant for such recommendation. We manually identified a number of properties after further interviews with participants. Relevant properties for Walter Cronkite (Wikidata entity *Q31073*) would be: country of citizenship, award received, date of birth, child, religion, occupation and date of death. These fields can be used to retrieve other stories that are linked through these properties and present them to the users as recommendations. There can be multiple strategies to ensure relevancy of these stories, for example by using semantic distance measures based on the number of matching property-values (a user can be recommended other stories mentioning US-born journalists born in the 1910s).

⁴ <https://www.beeldengeluid.nl/verhalen/hoe-cronkite-de-kijk-op-de-vietnamoorlog-veranderde>

4.5 Evaluation of internal cases

We evaluated the use cases and mockups in two evaluation rounds. In the first round, internal users of the R&D department were shown paper prototypes in short focus group sessions during which they were asked to reflect on the cases and prototypes. A proposed change for UC-I-1 was to further focus on people with a media background. For UC-I-2 the proposed change was to include possible archive material on the extra information page for a person mentioned in a story. It was also proposed to move extra information button from the top right to the bottom of the page as it could prove too distracting. UC-I-3 did not receive proposed changes at this stage..

The changes were implemented and presented for second round of evaluation in informal focus groups. These conversations allowed us to get in insight into the projected usefulness of the use cases. The updated UC-I-1 was evaluated as a useful tool for keeping track of copyright expirations. Further alignments between the GTAA and Wikidata are still needed to tackle the problem of incompleteness of the data. UC-I-2 and UC-I-3 were evaluated as interesting assets for the new story platform but further evaluations would need to be done on the basis of prototypes with higher fidelity.

5 Use Case for External Users

There are many external users of the archive, including creative (media) professionals, students, and interested laypeople. We here focus on (humanities) researchers that seek access to media resources, knowledge and use technical tools to support their research. Within this group, there are still different types of researchers with their own information needs. In the context of the CLARIAH project⁵, in which NISV is a partner, a digital research environment is being developed and this provided us with opportunities to engage with external users.

5.1 The CLARIAH Media Suite

The CLARIAH Media Suite [13] is an environment designed for media researchers and humanities scientists. It allows the user to explore, select and analyze audio-visual and textual material. It offers various tools to aid the user in their research. These tools allow the user to select the dataset that they are interested and carry out their research on it. In this research, we focus only on the “Search” Tool which allows the user to explore a dataset through textual keyword queries.

5.2 Elicitation method

To identify specifically relevant exploratory suggestions, five scholars were interviewed. The participants ranged from media studies scholars to digital historians, cultural heritage professionals and humanitarians. Each of these participants was a potential user of the Media Suite. They had different areas of research they focus on,

⁵ <http://www.clariah.nl>

specifically: 1) Drugs, 2) Sports, 3) Occupations, 4) History and 5) Disruptive media events. The semi-structured interview included questions about their area of expertise and type of research. The interviews aimed to find out how the participants used the CLARIAH Media Suite and to identify possible improvements. In order to give the participants an idea of the data a Wikidata page contains, several examples of Wikidata pages were shown. In order to see how Wikidata could play a role in the CLARIAH Media Suite, researchers were asked to share their thoughts.

5.3 UC-E-1: Exploratory extension of the CLARIAH Media Suite

The elicited use case focus on providing opportunities for advanced retrieval beyond this keyword search. Specifically, we focus supporting scholars in the *exploratory* research phase [14], where they do not yet know exactly the content and scope of the dataset nor their exact information need (no known item search).

In such an information retrieval scenario, researchers have a general idea of what they are interested in but do not know the specifics of their interest. For example, a researcher might be interested in all members who belong or have belonged to a certain political party. Such data is available in structured datasets (in our case in Wikidata) and therefore the proposed extension is to include additional results in an improved interface of the Media Suite.

5.4 Proof of concept

A proof-of-concept prototype was designed and implemented as an extension of the CLARIAH Media Suite. This “Wikidata retrieval service” is a look-up service that allows the user to set parameters to retrieve lists of matching *persons*. Figure 3 displays a prototype mock-up of the CLARIAH Media Suite with the additional functionality of Wikidata. The user can select a single property from this list and then search for entities belonging to that property.

Next to the retrieved person names, two buttons are depicted for GTAA and Wikidata, which redirect the user to the respective external web page. On mouseover, a small infobox with relevant information is displayed next to the cursor. The GTAA box contains information such as *skos:prefLabel*, *skos:scopeNote*, whereas the Wikidata box contains all known statements about the selected person (Figure 4).

Single Collection Search

The screenshot shows the 'Single Collection Search' interface. At the top, there are buttons for 'Set collection (Audiovisual Collection)' and 'Set project (Test)'. Below, the search is performed in the 'Audiovisual Collection' with a total of 1889624 records. The search term 'de vries' is entered in the search box, and the results are filtered to 1130033. A list of categories is shown, including 'voetbal', 'muziek', 'kinderen', etc. A pagination bar shows results 1 through 8. A specific result is highlighted: 'DE NIEUWS BV; De nieuws BV (2017 - 2018) (23-01-2018)'. To the right, the 'Wikidata Retrieval Service (?)' dropdown menu is open, listing various properties such as 'Award Received', 'Birth Name', 'Country of Citizenship', 'Date of Birth', 'Date of Death', 'Educated at', 'Gender', 'GTAA ID', 'Member of Political Party', 'Nominated for', 'Participant of', 'Place of Birth', 'Place of Death', '#plays Sport', 'Pseudonym', and 'Religion'. The 'Date of Death' and 'Place of Birth' fields are populated with values.

Fig. 3. Wikidata Retrieval Service - Dropdown Property List

The figure shows two side-by-side screenshots of the Wikidata Retrieval Service. The left screenshot shows the 'Wikidata Retrieval Service (?)' dropdown menu with a list of names and their corresponding Wikidata properties and values. The right screenshot shows the 'GTAA hover functions' for the same names, displaying additional information such as 'skos:prefLabel', 'skos:hiddenLabel', 'skos:notation', and 'skos:scopeNote'.

Fig. 4. Wikidata Retrieval Service: Wikidata (Left) and GTAA hover functions (Right)

This additional information assists the user in exploring possible follow-up queries around persons in the CLARIAH Media Suite Search Tool.

Finally, the user can select one or multiple persons by clicking on the desired person names. The selected persons are highlighted and once selected, are used as a filter for the CLARIAH Media Suite Search Tool. Once the selection has been confirmed, the user can use the existing search box to search for other terms and keywords.

This functionality allows the researcher in his or her exploratory phase to identify more specific information needs. The added functionality allows the user to retrieve a list of associated persons from Wikidata, and be able to filter by selecting the people who are displayed in the list. An additional benefit is to have a quick and easy overview of all facts belonging to a person. The Wikidata Retrieval Service uses SPARQL queries in the background to retrieve properties and values from Wikidata, but hides

the complexity of the query language for the user. The queries are generated on the fly according to the set parameters and sent live to the Wikidata Query Service⁶.

5.5 Qualitative evaluation

Setup. Several participants from the first interviews were asked to participate in an evaluation session to gather feedback and see if the proposed improvements add value. The participants were asked to sit in front of a computer which displayed the mock-up prototype and were asked to complete three tasks, which were based on realistic research tasks as identified in the earlier interview:

1. Assume you are a researcher interested in sports; in particular female sport participation in association football. Use the Wikidata Retrieval Service to retrieve a list of all female football players.
2. Assume you are a researcher interested in politics; in particular the political party “VVD”. Use the Wikidata Retrieval Service to retrieve a list of all members of the VVD who died before the year 2000.
3. Assume you are a researcher interested in disruptive media events; in particular, the death of Pim Fortuyn. You know that Pim Fortuyn died of a homicide, you decide to research whether there were any other related deaths by homicide during a time period of 2 years before and after his death (06-05-2002). Use the Wikidata Retrieval Service to retrieve a list of people who died of a homicide during a period of 2 years before and after Pim Fortuyn’s death. Unfortunately, you find no correlation between the retrieved persons. You decide to look up all people who died of the same cause of death: ballistic trauma. Use the Wikidata Retrieval Service to retrieve a list of people who died of a ballistic trauma.

After the participants had completed the tasks, they were asked questions relating to the Wikidata Retrieval Service about the added value to their research, what its drawbacks were and if they could think of any improvements.

Results. All three participants were able to complete the aforementioned tasks with relative ease. One stated that they felt it added value to a researcher, as researchers generally work in concepts and it is difficult to search for a list of individual persons. The participants raised questions of completeness of the data shown and stated that the data as of now is still lacking. User of the tool would have to be informed of the data incompleteness and be aware that the tool serves as an exploratory search, filling in missing knowledge. A suggested improvement would be to order the ontology in a useful way for researchers and to make the connections between entities clear; i.e.: where do they overlap? However, the participant also raised the issue of data completeness and questioned when it would be possible to say the data is complete enough to indeed add value. Another suggestion made by the participants is to cluster properties into categories in the user interface, as the current –complete- list of prop-

⁶ <https://query.wikidata.org/>

erties displayed was perceived to result in an information overload, causing the user to spend too much time searching for a specific property.

The third participant had an extensive knowledge of Wikidata and had worked on projects with Linked Open Data and the CLARIAH Media Suite. As such, they stated that integration with Wikidata can be very beneficial, especially for media researchers as it provides context for the user. This was seen as the most important enrichment as it serves as a support for exploration. Furthermore, they added that most users would not know about Wikidata's data (in)completeness and would have to be made aware of that before using the tool, perhaps in a tutorial beforehand.

6 Discussion

With the adoption of Linked Data by cultural heritage organizations, opportunities of reusing external general purpose datasets arise. However, in many cases, such opportunities stem from a 'technology push'. In this paper, we aim to identify these opportunities from the perspective of the heritage organization itself, more specifically from the perspective of internal and external users. We describe a user-centric method using interviews with different stakeholders to identify possibilities for using links from an organization-specific structured vocabulary to a general purpose knowledge base. The case study at NISV explores the possibilities that the links between GTAA and Wikidata provide for three types of internal users and for external users. The proofs of concepts indeed show added benefits to these users.

The method of iterative design of use cases and focus group discussions was shown to be a useful way of determining the added value of linked metadata. This research aimed to find answers to the questions of what use cases can be formed by an enriched GTAA and the differences in wishes between internal and external users are. This research also showed that when giving users insight in all the raw data and its properties, new ideas about the usefulness of said linked data can be identified. We show that focus on different properties within the data lead to completely different and independent use cases. The use cases that came forward in this research were connected to a specific focused group of properties. For UC-I-1 the focus lied on the "date of death" property. For UC-I-2 the focus lied on properties with a high occurrence rate. For UC-I-3 the focus lied on very specific properties that were handpicked for relevancy. For UC-E-1, all properties are used.

Differences between internal and external users include the access to and familiarity with backend tooling. Where to internal users, many tools are already available, for external users this might not be the case. There are differences in the wishes of the two groups of users. Internal users want to mainly use an enriched GTAA to set up further project or create new tools. Whereas the external users mainly want to use an enriched GTAA to receive as much extra information as possible as well as add some functionality to their search behavior. The similarities in the wishes mainly lie in the background information and added functionality to specific aspects of their workload.

A common comment in both user groups is that data completeness and quality are both crucial and unclear with respect to the external data. Even though Wikidata con-

tains a significant amount of information and it is being updated daily by an active community of ~19.000 users it is far from complete. Researchers cannot be assured of the retrieved data to be complete as there is always a possibility data is not available or incorrect. One way to mitigate this is to use the external data mostly to provide *extra* information and to use it in exploratory search activities.

Another potential downside of using external services such as the Wikidata SPARQL endpoint is that it relies on external servers and such servers might be down, unavailable or data might be changed without the user's knowing.

However, as a net result, both internal and external users are enthusiastic about the affordances that metadata and vocabularies linked to general purpose knowledge base give. Therefore, we expect further uptake of Linked Data at not only this organization, but others like it.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts pp. 205-227 (2009)
2. Hyvönen, Eero. "Publishing and using cultural heritage linked data on the semantic web." *Synthesis Lectures on the Semantic Web: Theory and Technology 2.1* (2012): 1-159.
3. Vrandečić & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78-85.
4. Erxleben, Fredo, et al. "Introducing Wikidata to the linked data web." *International Semantic Web Conference*. Springer, Cham, 2014.
5. Färber, Michael, et al. "Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago." *Semantic Web Preprint* (2016): 1-53.
6. Thornton, K., Cochrane, E., Ledoux, T., Caron, B., & Wilson, C. (2017). Modeling the Domain of Digital Preservation in Wikidata. IPress technical document. <https://ipres2017.jp/wp-content/uploads/7.pdf>
7. Debevere, Pedro, et al. "Linking thesauri to the Linked Open Data cloud for improved media retrieval." *12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-2011)*. 2011.
8. Tordai, A., Omelayenko, B., & Schreiber, G. (2007, October). Thesaurus and metadata alignment for a semantic e-culture application. In *Proceedings of the 4th international conference on Knowledge capture* (pp. 199-200). ACM.
9. Neubert, Joachim. "Bringing the" Thesaurus for Economics" on to the Web of Linked Data." *LDOW 25964* (2009).
10. Ahmeti, Albin, Simon Razniewski, and Axel Polleres. "Assessing the completeness of entities in knowledge bases." *European Semantic Web Conference*. Springer, Cham, 2017.
11. Brinkerink, Maarten. "Open Images: Towards an Audiovisual Commons." *OKCon*. 2011.
12. Maccatrozzo, Valentina. "Burst the filter bubble: using semantic web to enable serendipity." *International Semantic Web Conference*. Springer, Berlin, Heidelberg, 2012.
13. Carlos Martinez-Ortiz et al.: CLARIAH Media Suite: Enabling Scholarly Research for Distributed Audiovisual and Mixed Media Data Sets in a Sustainable Infrastructure. In: *proceedings of DHBenelux 2018*, Amsterdam, the Netherlands.
14. De Boer, et al. "DIVE into the event-based browsing of linked historical media." *Web Semantics: Science, Services and Agents on the World Wide Web* 35 (2015): 152-158.