

# Researching Wikidata's added value in accommodating audio-visual researchers' information needs

Case study at the Netherlands Institute for Sound & Vision

JOHN BROOKS, Vrije Universiteit, The Netherlands

## ABSTRACT

Linked Data is playing an increasingly significant role in the field of cultural heritage. This paper aims to find out how Linked Data techniques can add value for external researchers who collaborate at cultural heritage institutions and will take the form of a case study at the Netherlands Institute for Sound & Vision (NISV). NISV manages an audio-visual collection and is looking for ways to implement Linked Data principles on its collection. Specifically, NISV is looking at a general-purpose knowledge base, Wikidata<sup>1</sup>. This research is part of a broader project aiming to map the possibilities of Linked Data principles for internal and external audiences. This paper will focus on external audiences that collaborate at NISV and will describe these external researchers, including their information needs when it comes to accessing NISV's audio-visual collection. To see what information it contains and how this existing information matches external researchers' information needs, we analyse NISV's General Thesaurus for Audio-visual Archives (GTAA). We also analyse the Person facet of Wikidata to compare both datasets and see which data is relevant to align with the GTAA, specifically for the needs of external researchers. Based on this, we propose an improvement to the Clariah Mediasuite<sup>2</sup> as a mock-up prototype, which incorporates Wikidata as an additional functionality to serve as a method for performing exploratory research on external researchers' subjects of interest. A qualitative evaluation of the mock-up prototype by external researchers shows that the proposed addition to the Clariah Mediasuite adds value for external researchers, as long as they are carrying out research in an exploratory manner. We discuss limitations of the proposed addition, including incomplete and corrupted data, overabundance of properties, technical limitations and future work.

Additional Key Words and Phrases: Linked Data, Resource Description Framework (RDF), digital thesauri, cultural heritage

## 1 INTRODUCTION

### 1.1 The Netherlands Institute for Sound & Vision

The Netherlands Institute for Sound and Vision (NISV)<sup>3</sup> is an audio-visual archive located in Hilversum, the Netherlands. This cultural historic institute manages over 70% of all material relating to Dutch audio-visual heritage such as Dutch radio programs, television programs, documentaries and music.

Together with several other Dutch organisations who manage Dutch audio-visual heritage, NISV have developed the Common Thesaurus for Audio-visual Archives (GTAA)<sup>4</sup>. This thesaurus allows NISV to accurately characterise their audio-visual material. The GTAA incorporates a list of terms, which include personal 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. The thesaurus allows the user to connect these terms with audio-visual material and is used to describe and annotate their material.

With the help of linked data principles, connections can be made between the own collection and other data sources. 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.

Cross-linking between cultural heritage collections and Linked Data platforms allows institutes to enhance their

<sup>1</sup><https://www.wikidata.org/wiki/Wikidata:Introduction>

<sup>2</sup><https://www.clariah.nl/en/work-packages/focus-areas/media-studies>

Author's address: John Brooks, Vrije Universiteit, De Boelelaan 1105, Amsterdam, Noord Holland, 1081 HV, The Netherlands, j.a.brooks@student.vu.nl

<sup>3</sup><https://beeldengeluid.nl/en>

<sup>4</sup><http://gtaa.beeldengeluid.nl/>

collection with a vast amount of additional data, meta-data. NISV's goal is to use this meta-data to enrich their thesaurus for internal personnel as well as external audiences.

NISV attracts several external parties, such as private individuals, heritage professionals, media professionals, teachers, students and researchers, each with their own information needs.

In this paper, the focus will lie on external researchers who collaborate at NISV. External researchers seek access to media resources, knowledge and technical tools to support their research. One could say that NISV's media- and userdata serve as fuel for research in the sense that its infrastructure facilitates research by partners in science, heritage and industry. In order to understand the different types of external researchers, several interviews were held discussing their information needs when it comes to accessing NISV's audiovisual collection.

## 1.2 Wikidata

Vrandečić & Krötzsch (2014) state that Wikidata is continuous crowd-sourced knowledgebase where Wikipedia's data is cleaned and integrated to make it machine-readable as well as human-readable.

General-purpose knowledge bases such as Wikidata aim to collect as much factual information about the world as possible. [1] They store information about entities and information about relationships between these entities in the form of RDF (Resource Description Framework) triples. [2]

In essence, Wikidata allows triple stores to be queried with SPARQL to find the desired information.

Each term or object within Wikidata has its own label, aliases and description, including property-value pairs and qualifiers, which includes contextual information. These are confined within certain data types such as items, strings, date & time, geographical coordinates and URLs, which could be described as metadata. [1]

Furthermore, each object has a unique identifier, or URI, and several other identifiers from different sources, one being a GTAA identifier.

Thornton et al. (2017) state that this metadata is necessary for the identification and management of these

entities, as it allows digital preservation practitioners to automate interactions with these entities. [3]

Similarly, Gligorov et al. (2010) state that quality meta-data associated with collection items is prerequisite for successful information retrieval and collection management. [4] To reflect with NISV's media collection, Tordai et al., (2007) say that a typical collection consists of two parts, the object descriptions (metadata) and an in-house vocabulary (thesaurus). The authors use the SKOS core schema to represent a thesauri as SKOS provides a standard way to represent knowledge organization systems using RDF.[5]

As previously stated, data from Wikidata is available as RDF and can be queried [3] and thus makes alignment with Wikidata a valid option in the wish of converting thesauri to an open standard of indexing and retrieval of resources. [6]

However, Van Assem et al. (2006) argue that a standard schema for thesauri is required. In the paper, the authors present an analysis of the GTAA and its needs. For example, the GTAA has a subject facet and several other assets like alphabetically controlled lists and additional scope notes. [7]

The subject facet is similar to Wikidata's label facet which are both used to give a name to their entities. Furthermore, Wikidata's description facet is comparable to the GTAA scopeNote facet, which are used to describe the entity.

As NISV is looking to align the Person facet from the GTAA with Wikidata, these similarities can form the bridge between the two.

Wikidata recognises a GTAA ID as an external identifier. By looking for persons in the GTAA who match a Wikidata person page, it is possible to add the GTAA ID as an identifier to the corresponding Wikidata page.

NISV's goal is to automatically align as much entities from the person-axis of the GTAA with Wikidata. [14] However, there is still uncertainty on how this should be done and what benefits could be reaped from it.

*1.2.1 Wikidata Query Service.* Wikidata host the Wikidata Query Service<sup>5</sup> which allows a user to send SPARQL queries to the service to access Wikidata's content. The user is free to create their own SPARQL query or can

<sup>5</sup><https://query.wikidata.org/>

choose from a range of (adjustable) examples. As the knowledgebase is public domain<sup>6</sup>, all Wikidata's content can theoretically be retrieved by sending SPARQL queries to the service.

### 1.3 Research Questions

In order to help NISV to achieve their goal, we seek to answer the following research questions:

- How can Linked Open Data techniques play a role in NISV's audio-visual collection?
- Does the implementation of Linked Open Data techniques in NISV's audio-visual collection satisfy external researchers' information needs?

### 1.4 External Researchers

To answer these research questions, several researchers were contacted to participate in an interview. A total of five researchers from the following fields were interviewed:

- Drugs
- Sports
- Occupations
- History
- Disruptive Media Events

Several questions were asked during these interviews relating to their field of interest, specifically asking them to describe their area of expertise and what type of research they carry out. The participants ranged from media studies scholars to digital historians, cultural heritage professionals and humanitarians.

The interviews aimed to find out how the participants used the GTAA. Unfortunately, it became clear that almost none of the interviewees had any knowledge of the GTAA.

However, they did use the Clariah Mediasuite which uses the GTAA as a vocabulary to annotate audio-visual material. Instead, participants were asked to describe how they used the Clariah Mediasuite as of now and if, in their opinion, there is any way it can be improved.

Some suggestions were made, unfortunately none relating to Linked Open Data.

<sup>6</sup>[https://www.wikidata.org/wiki/Wikidata:SPARQL\\_query\\_service/A\\_gentle\\_introduction\\_to\\_the\\_Wikidata\\_Query\\_Service](https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/A_gentle_introduction_to_the_Wikidata_Query_Service)

Therefore, other questions were asked relating to their knowledge of Linked Open Data. All but one had any knowledge of the structured data method. This opened up the opportunity to show the participants the possibilities of Linked Open Data.

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 Mediasuite, researchers were asked to share their thoughts on the subject.

Furthermore, the participants were presented with a list of properties which can be found in section 3.1. The participants were asked to state which properties they deemed useful for their research.

Most researchers named the generic properties and some non-generic properties were also named by participants when it fit their research, i.e. someone interested in politics selected the properties relating to politics, e.g. 'member of political party'.

We found that researchers generally use the Clariah Mediasuite Search Tool and the Compare Tool to look up audio-visual material relating to the subject they are researching. Even though almost none of the researchers knew about the GTAA or Linked Data, the interviews still proved useful as the researchers had the opportunity to talk about the way they used the Clariah Mediasuite for their specific research.

In general, external researchers use the Clariah Mediasuite by searching for a term or multiple terms related to their research to view the list of results pertaining to the search query. The user then uses the provided filters to filter out non-relevant results.

From there on, the user selects audio-visual footage of interest and views it to see if it helps answer their research question.

### 1.5 Common Thesaurus for Audio-visual Archives (GTAA)

As stated before, the GTAA, the Common Thesaurus for Audio-visual Archives is a thesaurus that allows NISV to annotate its audio-visual material and is based

on the SKOS scheme<sup>7</sup>, a data model designed for the

<sup>7</sup><https://www.w3.org/2009/08/skos-reference/skos.html>

representation of thesauri. The GTAA contains several categories containing terms that can be used for annotation. A few example categories are: subjects, genres, cast members, speakers & personal names. However, this research focuses solely on the personal names category. Each entity in the personal name category in the GTAA has an identifier, the GTAA ID, and a preferred label, also known as `skos:prefLabel`, which is the most common known name for the person in question. Most entities also have a `skos:hiddenLabel`, which is a different name for the person in the question, often consisting of a different spelling. Furthermore, each entity has a `skos:scopeNote`, a short description of the corresponding person, usually consisting of the person's occupation and country of origin.

These are the four statements a single entity contains and thus the only point of information that can be relied on to determine whether the entity is indeed the entity a researcher is looking for.

Thus, this brings forward the problem that NISV is facing in their aim to automatically align the GTAA person axis with Wikidata, namely that this information is quite limited and mostly insufficient for accurate alignment. Unfortunately, the GTAA is mostly used internally at NISV as the interviews have shown external researchers generally have no knowledge of its existence.

## 1.6 Clariah Mediasuite

Instead, external researchers who collaborate at NISV use the Clariah Mediasuite for their research.

The Clariah Media Suite is an environment designed for media researchers and digital humanities scientists. The digital humanities community includes scholars with a wide diversity of research interests and goals; they all work with different types of data and their research objectives have specific requirements which cannot be easily facilitated by tools using a single, generic approach. [8]

The Clariah Mediasuite allows the user to select the dataset that they are interested in and to explore, select and analyse audio-visual and textual material. It is part of the CLARIAH infrastructure aiming to serve the needs of media scholars by providing access to audio-visual collections and their contextual data material. [8]

In alignment with this research, CLARIAH CORE developers are working on facilities for (mainly meta)data as Linked Data, the ability to search in the linked data and specifically Linking CMDI (Component Metadata) to Linked Data, and vice-versa. [11]

The Clariah Mediasuite offers the following four tools to aid the user in their research:

- Inspect
- Explore
- Search
- Compare

For its relevance to this project, we only discuss the Clariah Mediasuite Search Tool.

The Clariah Mediasuite Search tool allows the user to explore a collection or dataset by enabling the user to enter a search query. The user is able to select the dataset they want to research and then has the option to type in a search query.

The tool uses Elasticsearch, a distributed full-text search engine capable of addressing issues of scalability, big data search and performance [13], to search through its datasets, returning all relevant results to the specified query.

In addition, the user can specify where the tool should search for the specified query; field cluster search (e.g. all fields, titles, descriptions or subtitles). Furthermore, it allows the user to search for material between two dates, giving the user the ability to sort by date types ranging from creation date to production year. Once the results have been displayed, the user can filter by broadcaster, genre, keyword or create a new filter (faceted search). The user is able to browse through the collection's results and view its audio-visual material.

A more detailed description can be found on the Clariah Media Suite Documentation Page<sup>8</sup>

## 2 ANALYSIS OF THE DATA

### 2.1 Common Thesaurus for Audio-visual Archives (GTAA)

In order to understand the possibilities of Linked Open Data and to see how complete the data is, an analysis

<sup>8</sup><https://clariah.github.io/mediasuite-info/pages/tools.html>

Table 1. Occurrences of properties aligned persons GTAA & Wikidata

	Occurrences (OC)	% of total	Multiple entries (ME)	ME % of OC	ME % of total
Name	10,350	100.00%	0	0.00%	0.00%
Gender	10,294	99.46%	0	0.00%	0.00%
Birth name	839	8.11%	12	1.43%	0.12%
Pseudonym	124	1.20%	17	13.71%	0.16%
Date of birth	10,040	97.00%	63	0.63%	0.61%
Place of birth	7390	71.40%	41	0.55%	0.40%
Date of death	3644	35.21%	38	1.04%	0.37%
Place of death	2776	26.82%	17	0.61%	0.16%
Occupation	9988	96.50%	5581	55.88%	53.92%
Country of citizenship	9976	96.39%	493	4.94%	4.76%

of all persons aligned between the GTAA and Wikidata was made, i.e. all Wikidata entities containing a GTAA identifier.

The data in Table 1 was retrieved by sending SPARQL queries to the Wikidata Query Service<sup>9</sup> and is made up out of all persons who are aligned between the GTAA and Wikidata. The rows in Table 1 contains a list of properties belonging to a person. The columns contains the following information:

- The amount of times a property occurs in the alignments
- The amount of times a property occurs in the alignments / total amount of alignments
- The amount of times a property contained multiple entries
- The amount of times a property contained multiple entries / amount of occurrences for that property
- The amount of times a property contained multiple entries / total amount of alignments

As can be seen in Table 1, date of death and place of death stand out with a low occurrence rate. This can be attributed to most people in the dataset still being alive. Birth name and pseudonym also stand out with a low occurrence rate, however it is presumable that most people do not have a differing birth name or pseudonym.

In contrast, most generic properties such as name, gender, date of birth, place of birth have a high occurrence

<sup>9</sup><https://query.wikidata.org/>

rate ( $\geq 96.39\%$ ) This tells us that data completeness for generic properties is high but not yet whole and brings up the issue whether the data can be trusted upon. In accordance with Ahmeti et al. (2017), we found that generally, popular entries have a high completeness whereas less popular entries have a low completeness when looking at individual entities.[10] Table 1 gives us a small overview of the dataset we are working with. It is important to note that even though most properties have a high occurrence rate, they still do not have a complete occurrence rate of 100%. This tells us that data is still lacking or missing.

## 2.2 Wikidata

A similar analysis was made for the complete Person facet in Wikidata, i.e. all persons containing a GTAA identifier and all persons not containing a GTAA identifier. The results show us that whereas generic properties have a high occurrence rate, non-generic properties have a low occurrence rate. Also, one can see that Wikidata's properties in general have more English instances than the ones linked with the GTAA, which tend to be mostly Dutch.

## 2.3 Comparison of the datasets

The analysis of both datasets resulted in the following comparisons which show us a few differences between

the Wikidata entities with a GTAA identifier and Wikidata entities in general.<sup>10</sup>

For example, Wikidata entities with a GTAA identifier are mainly Dutch-oriented, with over half of all occurrences (51.22%) stating the Kingdom of the Netherlands as country of citizenship.

In contrast, when looking at country of citizenship in Wikidata as a whole, Kingdom of the Netherlands appears only 49,131 times, whereas the 3 most frequent countries of citizenship: USA, Germany & France, appear 371,640, 222,355 & 164,300 times respectively.<sup>11</sup> Furthermore, out of the 20 most frequent places of birth, there exist only 5 non-Dutch places.

However, in contrast, out of the 20 most frequent places of death, only 6 are Dutch places.

These high occurrences of Dutch citizens in the Wikidata entities with a GTAA identifier can be attributed to the fact that the GTAA is designed for annotation of Dutch audio-visual material.

Therefore, there is a high occurrence of Dutch people who have to do with Dutch television, radio, theater music or have appeared on these platforms.

We also find that the GTAA and Wikidata both contain a significant amount more males than females, where Wikidata as a whole contains 3,113,633 males and 692,064 females, giving us a 4.50 male to female ratio and where the Wikidata entities with a GTAA ID contains 8127 males and 2218 females, giving us a 3.66 male to female ratio.

More similarities exist, such as the fact that the most popular sport attributed to the 'plays sport' property is association football in both datasets, following by athletics.

However, as stated before, due to the GTAA being designed for Dutch annotation purposes, the 20 most frequent occupations, for a large part, exist of Dutch entities having to do with audio-visual material such as actors, presenters, screenwriters and directors.

<sup>10</sup>The complete results of the analysis of the GTAA and Wikidata can be found at [https://figshare.com/projects/Researching\\_Wikidata\\_s\\_added\\_value\\_in\\_accommodating\\_audio-visual\\_researchers\\_information\\_needs/36350](https://figshare.com/projects/Researching_Wikidata_s_added_value_in_accommodating_audio-visual_researchers_information_needs/36350)

<sup>11</sup>These results can be found in Appendix A & B

Whereas the 20 most frequent occupations of Wikidata entities as a whole consists of a mix of professions, ranging from journalist, to politicians and university teachers.

### 3 PROPOSED IMPROVEMENTS TO THE CLARIAH MEDIASUITE

#### 3.1 Property Selection

Wikidata has a finite amount of property labels; it would be possible to incorporate all Wikidata property labels into the system. However, we propose the following selection of properties relevant to external researcher's information needs:

- Awards Received
- Birth Name
- Cause of Death
- Country of Citizenship
- Date of Birth
- Date of Death
- Educated at
- Gender
- Manner of Death
- Member of Political Party
- Occupation
- Place of Birth
- Place of Death
- Plays Sport
- Pseudonym
- Nominated for
- Religion

These properties were selected on the basis of them being most relevant to external researchers as established in the interviews. The properties are divided into the following categories: generic, entertainment, sport, politics, religion & miscellaneous.

#### 3.2 The Clariah Mediasuite

To answer the question how Linked Open Data techniques can play a role in the Clariah Mediasuite, this section will describe proposed improvements to the Clariah Mediasuite.

As stated before, the Clariah Mediasuite offers a search tool which allows the user to browse through a collection or dataset of the user's choice.

The user can enter a query in the search tool to retrieve a list of all audio-visual material that has to do with the entered query. The tool allows the user to view this audio-visual material, if available and it enables the user to enter different types of queries, which can range from subjects to concepts and personal names.

## Single Collection Search

The screenshot displays a search interface for the 'Audiovisual Collection'. At the top, there are buttons for 'Set collection (Audiovisual Collection)' and 'Set project (Test)'. Below this, the search context is 'Searching in : Audiovisual Collection' with a total of 1889624 records. A search box contains the text 'de vries'. The results are filtered to 1130033 items. A list of filters is shown: Broadcaster, Genre (series), Keyword, and NEW. A list of categories with counts is provided: voetbal (12010), muziek (11239), kinderen (9469), betogingen (8916), popmuziek (8333), tentoonstellingen (5291), politiek (5139), toneel (5139), geschiedenis (5061), zangers (487), vluchtelingen (4605), and poëzie (4541). A pagination bar shows pages 1 through 8. A search result for 'DE NIEUWS BV; De nieuws BV (2017 - 2018)' is visible. On 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 'Retrieve by property:' dropdown is set to 'Last'.

Fig. 1. Wikidata Retrieval Service - Dropdown Property List

The previously mentioned interviews with external researchers have shown that most researchers who use the search tool have a specific goal in mind. They have a thorough understanding of the subject they are interested in and know how to formulate a search query that fits their needs.

The search tool as-is aids a researcher with sufficient knowledge of a subject to carry out their research. However, there exist researchers who have just started out researching a certain subject. These researchers have a general idea what they are interested in but do not know the specifics of their interest, i.e. researchers who are researching in an exploratory manner. For example, a researcher might be interested in all members who belong or have belonged to a certain political party. Although this data is freely available on the Internet, the Clariah Mediasuite does not, as of yet, allow the user to look up this information when using their tools. Wikidata, however, does allow this information to be retrieved by sending a SPARQL query to their server. Based on this, we propose the following improvements

to the Clariah Mediasuite which incorporate the functionalities of Wikidata.

### 3.3 The Wikidata Retrieval Service

The Wikidata Retrieval Service can be described as a look-up service that allows the user to set parameters to retrieve lists of persons matching these parameters.

Figure 1 displays a prototype mock-up of the CLARIAH Mediasuite with the additional functionality of Wikidata.

The design of the tool is almost the same, except for the addition of the Wikidata Retrieval Service. The user selects the drop-down button to display all selected Wikidata properties and can use this function to search by a property of their choice. The Dutch political party GreenLeft will be used in the following example.

## Single Collection Search

Set collection (Audiovisual Collection) Set project (Test)

Searching in : Audiovisual Collection  
Total amount of records in this collection: 1889624

de vries

Total number of results based on "de vries" and selected filters: 1130033

Select date field Start date

Broadcaster ✕ Genre (series) ✕ Keyword ✕ NEW +

voetbal (12010) muziek (11239) kinderen (9469) betogingen (8916) popmuziek (8333) tentoonstellingen (5291) politiek (5139) toneel (5139) geschiedenis (5061) zangers (487) vluchtelingen (4605) poëzie (4541)

Wikidata Retrieval Service (?)

GreenLeft (Groenlinks)

Hafkamp, Hetty	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Willems, Wilbert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Voortman, Linda	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dibi, Tefik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gent, Ineke van	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Broekhoven, Karel van	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Barske, Anton	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lucas, Marjan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brand, Arie van den	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
more				

35 burgeroorlogen (6375) politiek (5634) 786 televisieprogramma's (4702) vrouwen (4621)

1 2 3 4 5 6 7 8 9 10 11 Next Last

DE NIEUWS BV; De nieuws BV (2017 - 2018) (23-01-2018)

Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke werkdag van 12.00 uur tot 14.00 uur met Willemijn Veenhoven en Patrick Lodiers.

Fig. 2. Wikidata Retrieval Service - Person Retrieval

The user can select a single property from this list and then search for entities carrying that property

In figure 2, the user has searched for the Dutch political party Greenleft (Groenlinks). As can be seen, a list has been returned of all members belonging to the chosen political party.

The retrieved list of persons takes the GTAA skos:prefLabel as input for displaying the names.

Next to the retrieved personal names, 2 buttons are depicted; on the left: the GTAA logo, on the right: the Wikidata logo. When clicked on, the user will be directed to the corresponding web page. When the mouse is hovered over it, a small box containing the contents of the Wikidata page or the GTAA page will appear.

The GTAA box contains the relevant GTAA information such as the skos:prefLabel, skos:hiddenLabel and the skos:scopeNote (right), whereas the Wikidata box contains all known statements about the selected person (left), as can be seen in figure 3.

This additional information could help the user in making the choice which specific persons they want to research, i.e. which persons they want to use in their search query in the existing Clariah Mediasuite Search Tool.

Finally, the user can select one or multiple persons by clicking on the desired personal names. The selected persons are highlighted and once selected, are used as a filter for the CLARIAH Mediasuite Search Tool. Once the selection has been confirmed, the user can use the existing search box to search for other terms and keywords.<sup>12</sup>

The idea behind the added functionality is in the fact that a researcher might not know the specific details of what they are trying to research. A researcher might be

<sup>12</sup>The complete list of figures can be found in Appendix C and the mock-up prototype can be found at [https://figshare.com/projects/Researching\\_Wikidata\\_s\\_added\\_value\\_in\\_accommodating\\_audio-visual\\_researchers\\_information\\_needs/36350](https://figshare.com/projects/Researching_Wikidata_s_added_value_in_accommodating_audio-visual_researchers_information_needs/36350)



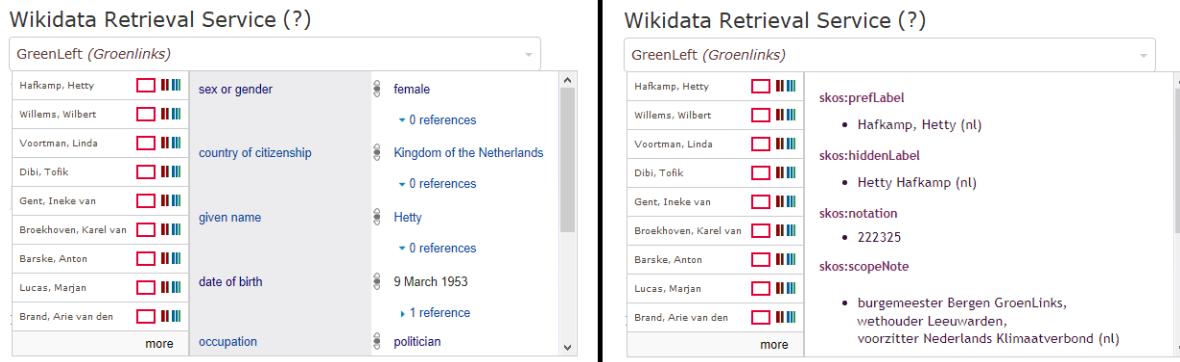


Fig. 3. Wikidata Retrieval Service - Wikidata Mouse-over Function (Left) | GTAA Mouse-over Function (Right)

researching a specific political party, occupation, religion, etc. but might not know the specific people associated with that property. 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 by using the Wikidata or GTAA button once the list of persons has been retrieved. Theoretically, as much properties as the user wishes to use can be set. However, this will naturally lead to 0 results quickly. The main point lies in the fact that multiple properties can be combined and set to retrieve persons who adhere to these multiple set properties.

The Wikidata Retrieval Service's functioning relies on Wikidata and its Wikidata Query Service. For example, when the property has been selected and the search bar is being used, a SPARQL query is sent to the Wikidata Query Service. This SPARQL query retrieves all related instances and orders them in a list. Once (part of) the query has been typed, the corresponding match is selected and displayed to the user. The user is then free to select the relevant results.

Furthermore, once multiple properties has been set and the user wishes to view the result, another SPARQL query is sent to the Wikidata Retrieval Service containing all set parameters.

Generally speaking, the Wikidata Retrieval Service works as an interface for users to send SPARQL queries to the Wikidata Query Service without having any prior

knowledge of SPARQL. SPARQL queries are generated on the fly according to the set parameters and sent live to the Wikidata Query Service, allowing the user to make use of the Wikidata Query Service without any prior knowledge of SPARQL. Generation of the appropriate SPARQL query happens behind the scenes, unbeknownst to the user. The user only gets to see the final product, i.e. the retrieved list.

### 3.4 Qualitative evaluation

In order to answer the question whether the implementation of Linked Open Data techniques in the Clariah Mediasuite satisfies external researchers' information needs, a working mock-up prototype of the Wikidata Retrieval Service was created.

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 to their research and increase ease of use when using the Clariah Mediasuite. The participants were asked to sit in front of a computer which displayed the mock-up prototype and were asked to complete four tasks.

The four tasks are as follows:

- 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.
- Assume you are a researcher interested in politics, in particular the political party VVD (People's Party for Freedom & Democracy). Use the Wikidata Retrieval Service to retrieve a

list of all members of the VVD who died before the year 2000.

- Assume you are a researcher interested in disruptive media events, in particular the death of Pim Fortuyn, a Dutch politician. 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 need decide to look up all people who died of the same cause of death: a gunshot wound, also known as a 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 a few questions relating to the Wikidata Retrieval Service.

Specifically they were asked if they thought it added value to their research, what its drawbacks were and if they could think of any improvements. Furthermore, they were presented with the following SPARQL query, which is used behind the scenes to retrieve the results in the third task:

```
SELECT ?q ?GTAA_ID ?qLabel ?date_of_death
WHERE {
  ?q wdt:P1741 ?GTAA_ID.
  ?q wdt:P1196 wd:Q149086.
  ?q wdt:P570 ?date_of_death
FILTER
(?date_of_death >= "2000-06-05T00:00:00Z"
^^xsd:dateTime &&
?date_of_death<= "2004-06-05T00:00:00Z"
^^xsd:dateTime) .
  SERVICE wikibase:label { bd:serviceParam
    wikibase:language "[AUTO_LANGUAGE]". }
}
```

The participants were then asked questions relating to their comprehension of it and general researcher comprehension of SPARQL.

### 3.5 Results

Of all three participants, all were able to complete all four tasks successfully in a period of 15 minutes, despite some minor inconsistencies with the prototype itself. One stated that they felt it added value to a researcher, as researchers generally work in concepts and it is quite difficult to search for a list of individual persons online. They also raised questions of completeness and stated that the data as of now is still lacking. Users of the tool would have to be informed of the data incompleteness and be aware that the tool serves as an exploratory search to fill in missing knowledge.

A suggested improvement would be make the connections between entities clear; i.e.: where do they overlap? This user had a limited knowledge of SPARQL and was partially able to comprehend the aforementioned query, stating that the problem with SPARQL is that the user needs to know not only the syntax, but also the identifiers, relations and datasets.

The second participant however, had no knowledge of SPARQL and thus was not able to understand the query at all. They agreed that behind-the-scenes SPARQL query generation could be beneficial to researchers and improved ease of use.

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 comment made would be to divide the properties into genres or categories, as the display of the properties resulted 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 already worked on projects with Linked Open Data and the Clariah Mediasuite.

As such, they stated that integration with Wikidata can be very beneficial, especially for media-researchers as it creates a feeling of context for the user. This gaining of insight into a subject 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 incompleteness and would have to be made aware of that before using the tool,

possibly in a tutorial beforehand.

The before-mentioned issues and suggested improvements are discussed in the following section.

## 4 DISCUSSION

### 4.1 Incomplete or corrupt data

Even though Wikidata contains a significant amount of information and it is being updated daily by an active community of 19.000 users<sup>13</sup>, it is far from complete. Researchers cannot rely on the retrieved data to be complete as there is always a possibility it is lacking. Furthermore, as the Wikidata community is crowd-sourced there is also the possibility that mistakes have been made when entering the data. Therefore, the retrieved data cannot be presented as fact and the issue of data completeness, a data quality measure that refers to the degree to which all required information is present in a particular dataset arises. [12]

Razniewski et. al (2016) rank the quality of a knowledge-base such as Wikidata on three dimensions:

- Size - Today's knowledge-bases can contain millions, if not billions of triples
- Precision - the proportion of triples that are correct
- Recall - the proportion of facts of the real world that are covered by the knowledge-base

and found that, similar to previous research, knowledge-bases do not score high on these three criteria. They also state the knowledge-bases suffer multiple problems, such as missing triples and the fact that they do not know how many triples are missing, or whether some are missing at all. [2]

Instead, the Wikidata Retrieval Service serves as a supportive exploratory tool to gain an insight into certain subjects.

### 4.2 Overabundance of properties

The mock-up prototype displays a list of all selected properties. This list contains only a selection of properties which we deemed relevant for external researchers. However, Wikidata offers many more properties and assumingly, these properties could also be relevant for external researchers. Therefore, this brings up the issue

on how to order the properties. Design-wise, it is unrealistic to have all properties displayed in one list. For inclusion of all properties Wikidata has to offer, we propose categorising each property in a certain classes, possibly with subclasses, such as generic, politics-related, sports-related, etc.

### 4.3 Technical limitations

The Wikidata Retrieval Service relies on Wikidata's SPARQL endpoint. Once too many properties have been set or these properties are too complicated, the amount of necessary processing power becomes too high and the Wikidata Query Service will time out. As the system retrieves data live from the Wikidata SPARQL endpoint, too complex queries will take too long and be unable to complete. In fact, an experiment by Buil-Aranda et al. (2013), where the following query (a query that should be inexpensive to compute):

```
SELECT * WHERE { ?s ?p ?o } LIMIT 100002
```

was sent to several SPARQL endpoints. The results showed that 233 endpoints (54.6%) returned no results. In contrast, only 57 endpoints (13.3%) returned 100,002 results. [15] This shows us that the vast majority of SPARQL endpoints is not able to handle large SPARQL queries.

### 4.4 Broader Project

As mentioned before, this research is part of a broader project. This case study was a collaborative effort by Tim de Bruyn and myself at NISV, where the main difference lies in the target audience.

Tim de Bruyn's research focuses on internal staff at NISV's information needs whereas this research focuses on external researchers. NISV's goal is sustainably link the terms from the GTAA to an external open knowledge base, i.e. Wikidata with the aim to facilitate links between their collection and enrichments (additional data and / or context) from other collections and databases.

The original set-up was to enrich the thesaurus (GTAA) with Wikidata items, however, due to the lack of knowledge the interviewed researchers had of the GTAA, the project changed course midway. Instead, this research focuses on enriching the Clariah Mediasuite with the

<sup>13</sup><https://www.wikidata.org/wiki/Wikidata:Statistics>

possibilities of Wikidata, thus creating the Wikidata Retrieval Service. As stated before, the main difference between this research and Tim de Bruyn's research lies in the target audience. However, both researches help NISV by showing the possibilities of Wikidata and in what way these could be implemented in order to further digitize their audio-visual collection.

#### 4.5 Generalisability

This research aimed to help NISV in finding out how meta-data from an online knowledge bases, Wikidata, can help enhance their audio-visual collection and thus help improve ease of use for external researchers who collaborate at NISV.

Specifically, this was a case study at NISV using data from its collection, data from a specific knowledge-base and with specific external end-users and specific information needs.

However, we feel that the proposed methods could be beneficial to all types of institutions who are looking for ways to enhance their collections. Using Linked Open Data techniques opens up the possibility to enhance many collections who incorporate a list of terms, such as geographical locations, persons, company names and so on, as most of these datasets contain a limited amount of information.

Most importantly, the proposed techniques are not limited to digital thesauri, as using meta-data from knowledge bases can be beneficial to all types of knowledge-sharing facilities.

Furthermore, there exist many structured data knowledgebases such as Wikidata, including DBPedia<sup>14</sup>, & GeoNames<sup>15</sup>, which can be used to enhance ease of use and user experience when accessing datasets. Lastly, this research focused on external researchers at NISV but the proposed methods could be useful to any person making use of a provided dataset to gain context and insights into a subject.

#### 4.6 Future Work

Many research has been done on knowledge-bases and their (in)completeness [2] [3] [10], stating issues such

<sup>14</sup><https://wiki.dbpedia.org/>

<sup>15</sup><http://www.geonames.org/>

missing and false statements. However, none of this research concludes when a knowledge-base's data can be deemed complete enough to be used as fact in scientific research. The issue remains in the fact that is practically impossible for a knowledge-base to be complete on all property statements it contains.

Furthermore, as stated before, this research focuses on specific end-users and specific information needs. However, future work may be done applying the methods discussed in this paper for different end-users with differing information needs, in order to find out if differing information needs require a variation of the proposed methods.

## 5 CONCLUSION

This paper describes a case study at the Netherlands Institute for Sound & Vision, where the aim was to enrich their thesaurus, the Common Thesaurus for Audio-Visual Archives (GTAA) with meta-data from an open knowledge base, i.e. Wikidata, to enhance its use for external researchers. This research aimed to answer the following research questions:

- How can Linked Open Data techniques play a role in NISV's audio-visual collection?
- Does the implementation of Linked Open Data techniques in NISV's audio-visual collection satisfy external researchers' information needs?

In order to answer these research questions interviews with external researchers were held and it became clear that researchers generally had no knowledge of the GTAA.

Therefore the scope changed and the focus changed to enriching the Clariah Mediasuite, a digital research environment designed for digital humanities scientists and media scholars to use to select, explore and analyse datasets containing audio-visual material.

Furthermore, an analysis of the data within the GTAA had to be made as well as the data within Wikidata.

To gain an oversight of the data and to see how Linked Open Data techniques could play a role for external researchers at NISV, a comparison of these datasets is made describing the similarities and differences between them, such as a similar male-to-female ratio in both datasets and differences in the most frequent entities per dataset.

From thereon, we propose an addition to the Clariah Mediasuite, The Wikidata Retrieval Service, which incorporates the functionalities of Wikidata. The Wikidata Retrieval Service allows the user to set certain parameters to retrieve a list of person names from Wikidata, i.e. search by properties to retrieve a list of person names who match the set properties. Furthermore, the Wikidata Retrieval Service allows the user to view additional information about the retrieved persons, i.e. meta-data, containing all known information from Wikidata and the GTAA. We create a working mock-up prototype of the Clariah Mediasuite with the additional functionality of Wikidata, the Wikidata Retrieval Service. In order to see whether the implementation of Linked Open Data techniques in NISV's audio-visual collection satisfies external researchers' information needs, we ask several researchers to perform four tasks using the mock-up prototype and ask them to evaluate it.

The evaluation sessions conclude that the Wikidata Retrieval Service is indeed beneficial to external researchers, however only in an exploratory fashion. Issues of data completeness are raised as Wikidata's data cannot be presented as complete or as fact. Instead, the Wikidata Retrieval Service allows the user to gain insight into persons relating to the subject they are researching and create context.

Limitations are discussed, such as the technical limitations of Wikidata's SPARQL endpoint, which cannot handle complex queries. Furthermore, issues that were raised in the evaluation sessions are discussed, such as the issue of incomplete or corrupt data and the ordering of the properties within the Wikidata Retrieval Service. In conclusion, this paper present a way to incorporate Wikidata on a specific research environment, the Clariah Mediasuite. However, the paper also highlights the fact that Wikidata can be beneficial for several audiences. Furthermore, it shows that the possibilities of Wikidata are not limited to the Clariah Mediasuite, but can be implemented on audio-visual research environments and prove to be useful for the end-user, albeit in an exploratory manner.

## REFERENCES

- [1] Vrandečić & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78-85.
- [2] Razniewski, S., Suchanek, F., & Nutt, W. (2016). But What Do We Actually Know?. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction* (pp. 40-44).
- [3] Thornton, K., Cochrane, E., Ledoux, T., Caron, B., & Wilson, C. (2017). Modeling the Domain of Digital Preservation in Wikidata
- [4] Gligorov, R., Baltussen, L. B., van Ossenbruggen, J., Aroyo, L., Brinkerink, M., Oomen, J., & van Ees, A. (2010). Towards integration of end-user tags with professional annotations
- [5] Tordai, A., Omelayenko, B., & Schreiber, G. (2007, October). The-saurus and metadata alignment for a semantic e-culture application. In *Proceedings of the 4th international conference on Knowledge capture* (pp. 199-200). ACM.
- [6] Fugazza, C., & Luraschi, G. (2012). Semantics-Aware Indexing of Geospatial Resources Based on Multilingual Thesauri: Methodology and Preliminary Results. *IJSDIR*, 7, 16-37.
- [7] Van Assem, M., Malaisé, V., Miles, A., & Schreiber, G. (2006, June). A method to convert thesauri to SKOS. In *European Semantic Web Conference* (pp. 95-109). Springer, Berlin, Heidelberg
- [8] Martinez-Ortiz, C., Ordelman, R., Koolen, M., Noordegraaf, J. J., Melgar, L., Aroyo, L., ... & Kiewik, R. (2017, July). From Tools to "Recipes": Building a Media Suite within the Dutch Digital Humanities Infrastructure CLARIAH.
- [9] Polleres, A., Feier, C., & Harth, A. (2006, June). Rules with contextually scoped negation. In *European Semantic Web Conference* (pp. 332-347). Springer, Berlin, Heidelberg.
- [10] Ahmeti, A., Razniewski, S., & Polleres, A. (2017, May). Assessing the completeness of entities in knowledge bases. In *European Semantic Web Conference* (pp. 7-11). Springer, Cham.
- [11] Odijk, J. (2016). CLARIAH in the Netherlands. In *LREC*.
- [12] Zaveri, A., Rula, A., Maurino, A., Pietrobbon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, 7(1), 63-93.
- [13] Kononenko, O., Baysal, O., Holmes, R., & Godfrey, M. W. (2014, May). Mining modern repositories with elasticsearch. In *Proceedings of the 11th Working Conference on Mining Software Repositories* (pp. 328-331). ACM.
- [14] Nederlands Instituut voor Beeld en Geluid, "Beleidsplan 2016 - 2020 Nederlands Instituut voor Beeld en Geluid: Naar een multimediale toekomst in beeld en geluid." , 21.
- [15] Buil-Aranda, C., Hogan, A., Umbrich, J., & Vandenbussche, P. Y. (2013, October). Sparql web-querying infrastructure: Ready for action?. In *International Semantic Web Conference* (pp. 277-293). Springer, Berlin, Heidelberg.

A APPENDIX - GTAA STATISTICS

Table 2. Gender - most frequent GTAA entities

<b>Gender</b>	<b>Count</b>	<b>Percentage</b>	<b>Rank</b>
Male	8127	78,54%	1
Female	2218	21,43%	2
Other	4	0.03%	3
<b>Total</b>	<b>10,348</b>	<b>100.00%</b>	

Table 3. Country of Citizenship - 10 most frequent GTAA entities

<b>Country of Citizenship</b>	<b>Count</b>	<b>Percentage</b>	<b>Rank</b>
Kingdom of the Netherlands	5475	51.22%	1
United States of America	1083	10.13%	2
United Kingdom	526	4.92%	3
Germany	392	3.67%	4
France	370	3.46%	5
Belgium	328	3.07%	6
Italy	212	1.98%	7
Soviet Union	125	1.17%	8
Austria	80	0.75%	9
Spain	78	0.73%	10
<b>Total</b>	<b>10,690</b>	<b>100.00%</b>	

Table 4. Occupation - 10 most frequent GTAA entities

<b>Occupation</b>	<b>Count</b>	<b>Percentage</b>	<b>Rank</b>
politician	1722	7.59%	1
writer	1344	5.93%	2
actor	1206	5.32%	3
journalist	976	4.30%	4
singer	780	3.44%	5
television presenter	759	3.35%	6
association football player	631	2.78%	7
film actor	560	2.47%	8
university teacher	538	2.37%	9
composer	516	2.28%	10
<b>Total</b>	<b>22,685</b>	<b>100.00%</b>	

## B APPENDIX - WIKIDATA STATISTICS

Table 5. Gender - most frequent Wikidata entities

Gender	Count	Percentage	Rank
Male	3,113,633	81.78%	1
Female	692,064	18.18%	2
Other	828	0,02%	3
Total	3,807,526	100,00%	

Table 6. Country of Citizenship - 10 most frequent Wikidata entities

Country of Citizenship	Count	Percentage	Rank
United States of America	371,640	51,22%	1
Germany	222,355	10,13%	2
France	164,300	4,92%	3
Ming dynasty	124,808	3,67%	4
United Kingdom	124,105	3,46%	5
Japan	116,664	3,07%	6
Spain	93,988	1,98%	7
Italy	80,517	1,17%	8
Soviet Union	72,794	0,75%	9
Qing dynasty	64,478	0,73%	10
Total	2,969,784	100,00%	

Table 7. Occupation - 10 most frequent Wikidata entities

Occupation	Count	Percentage	Rank
politician	414,082	10.03%	1
association football player	232399	5,63%	2
actor	184,586	4.47%	3
writer	176,888	4.29%	4
researcher	148,354	3.59%	5
painter	126,932	3.08%	6
journalist	102,658	2.49%	7
university teacher	80,599	1.95%	8
singer	69,122	1.67%	9
poet	65,657	1.59%	10
Total	4,126,998	100,00%	

## C APPENDIX - THE WIKIDATA RETRIEVAL SERVICE

### Single Collection Search

Set collection (Audiovisual Collection)
Set project (Test)

Searching in : **Audiovisual Collection**  
 Total amount of records in this collection: **1889624**

Total number of results based on "de vries" and selected filters: **1130033**

Select date field

Start date

Broadcaster ✕
Genre (series) ✕
Keyword ✕
NEW +

voetbal (12010) muziek (11239) kinderen (9469) betogingen (8916) popmuziek (8333) tentoonstellingen (5291) politie (5139) toneel (5139) geschiedenis (5061) zangers (487) vluchtelingen (4605) poëzie (4541)

1

2

3

4

5

6

7

8

**DE NIEUWS BV; De nieuws BV (2017 - 2018) (23-01-2018)**  
 Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke vrijdag met Patrick Lodiers.

Wikidata Retrieval Service(?)

Retrieve by property:

Award Received	
Birth Name	
Country of Citizenship	
Date of Birth	
Date of Death	e <input type="text"/>
Educated at	
Gender	
GTAA ID	35) burgeroorlogen (6375) politiek (5634)
Member of Political Party	786) televisieprogramma's (4702) vrouwen (4621)
Nominated for	
Participant of	
Place of Birth	Last
Place of Death	
Plays Sport	
Pseudonym	14.00 uur met Willemijn Veenhoven en Patrick
Religion	

Fig. 4. Wikidata Retrieval Service - List of properties



## Single Collection Search

Set collection (Audiovisual Collection) Set project (Test)

Searching in : Audiovisual Collection  
Total amount of records in this collection: 1889624

de vries] Q

Wikidata Retrieval Service (?)  
GroenLeft (GroenLinks)  
green political party operating in the Netherlands

Total number of results based on "de vries" and selected filters: 1130033

Select date field Start date End date

Broadcaster x Genre (series) x Keyword x NEW +

voetbal (12010) muziek (11239) kinderen (9469) betogingen (8916) popmuziek (8333) jeugd (8153) verkiezingen (7235) burgeroorlogen (6375) politiek (5634) tentoonstellingen (5291) politie (5139) toneel (5139) geschiedenis (5061) zangers (4870) sport (4845) terrorisme (4786) televisieprogramma's (4702) vrouwen (4621) vluchtelingen (4605) poëzie (4541)

1 2 3 4 5 6 7 8 9 10 11 Next Last

DE NIEUWS BV; De nieuws BV (2017 - 2018) (23-01-2018)  
Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke werkdag van 12.00 uur tot 14.00 uur met Willemijn Veenhoven en Patrick Lodiers.

Fig. 5. Wikidata Retrieval Service - Search by property

## Single Collection Search

Set collection (Audiovisual Collection) Set project (Test)

Searching in : Audiovisual Collection  
Total amount of records in this collection: 1889624

de vries] Q

Wikidata Retrieval Service (?)  
GreenLeft (Groenlinks)  
Halkamp, Hetty  
Willems, Wilbert  
Voortman, Linda  
Dibi, Tofik  
Gent, Ineke van  
Broekhoven, Karel van  
Barske, Anton  
Lucas, Marjan  
Brand, Arie van den  
more

Total number of results based on "de vries" and selected filters: 1130033

Select date field Start date End date

Broadcaster x Genre (series) x Keyword x NEW +

voetbal (12010) muziek (11239) kinderen (9469) betogingen (8916) popmuziek (8333) jeugd (8153) verkiezingen (7235) burgeroorlogen (6375) politiek (5634) tentoonstellingen (5291) politie (5139) toneel (5139) geschiedenis (5061) zangers (4870) sport (4845) terrorisme (4786) televisieprogramma's (4702) vrouwen (4621) vluchtelingen (4605) poëzie (4541)

1 2 3 4 5 6 7 8 9 10 11 Next Last

DE NIEUWS BV; De nieuws BV (2017 - 2018) (23-01-2018)  
Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke werkdag van 12.00 uur tot 14.00 uur met Willemijn Veenhoven en Patrick Lodiers.

Fig. 6. Wikidata Retrieval Service - Retrieved List

## Single Collection Search

[Set collection \(Audiovisual Collection\)](#)
[Set project \(Test\)](#)

**Searching in : Audiovisual Collection**  
 Total amount of records in this collection: 1889624

Total number of results based on "de vries" and selected filters: **1130033**

Select date field

[Broadcaster](#) ✕ [Genre \(series\)](#) ✕ [Keyword](#) ✕ [NEW](#) +

[voetbal \(12010\)](#) [muziek \(11239\)](#) [kinderen \(9469\)](#) [betogingen \(8916\)](#) [popmuziek \(8333\)](#)  
[tentoonstellingen \(5291\)](#) [politie \(5139\)](#) [toneel \(5139\)](#) [geschiedenis \(5061\)](#) [zangers \(487\)](#)  
[vluchtelingen \(4605\)](#) [poëzie \(4541\)](#)

Wikidata Retrieval Service (?)

GreenLeft (*Groenlinks*)

Hafkamp, Hetty	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	sex or gender	female
Willems, Wilbert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	country of citizenship	Kingdom of the Netherlands
Voortman, Linda	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	given name	Hetty
Dibi, Tofik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	date of birth	9 March 1953
Gent, Ineke van	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	occupation	politician
Broekhoven, Karel van	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Barske, Anton	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Lucas, Marjan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Brand, Arie van den	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
<a href="#">more</a>					

Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke werkdag van 12.00 uur tot 14.00 uur met Willemijn Veenhoven en Patrick Lodiers.

Fig. 7. Wikidata Retrieval Service - Wikidata Mouse Hover Function

## Single Collection Search

[Set collection \(Audiovisual Collection\)](#)
[Set project \(Test\)](#)

**Searching in : Audiovisual Collection**  
 Total amount of records in this collection: 1889624

Total number of results based on "de vries" and selected filters: **1130033**

Select date field

[Broadcaster](#) ✕ [Genre \(series\)](#) ✕ [Keyword](#) ✕ [NEW](#) +

[voetbal \(12010\)](#) [muziek \(11239\)](#) [kinderen \(9469\)](#) [betogingen \(8916\)](#) [popmuziek \(8333\)](#)  
[tentoonstellingen \(5291\)](#) [politie \(5139\)](#) [toneel \(5139\)](#) [geschiedenis \(5061\)](#) [zangers \(487\)](#)  
[vluchtelingen \(4605\)](#) [poëzie \(4541\)](#)

Wikidata Retrieval Service (?)

GreenLeft (*Groenlinks*)

Hafkamp, Hetty	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	skos:prefLabel	<ul style="list-style-type: none"> <li>Hafkamp, Hetty (nl)</li> </ul>
Willems, Wilbert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	skos:hiddenLabel	<ul style="list-style-type: none"> <li>Hetty Hafkamp (nl)</li> </ul>
Voortman, Linda	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	skos:notation	<ul style="list-style-type: none"> <li>222325</li> </ul>
Dibi, Tofik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	skos:scopeNote	<ul style="list-style-type: none"> <li>burgemeester Bergen GroenLinks, vethouder Leeuwarden, voorzitter Nederlands Klimaatverbond (nl)</li> </ul>
Gent, Ineke van	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Broekhoven, Karel van	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Barske, Anton	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Lucas, Marjan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Brand, Arie van den	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
<a href="#">more</a>					

Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke werkdag van 12.00 uur tot 14.00 uur met Willemijn Veenhoven en Patrick Lodiers.

Fig. 8. Wikidata Retrieval Service - GTAA Mouse-Over Function

## Single Collection Search

Set collection (Audiovisual Collection) Set project (Test)

Searching in : Audiovisual Collection  
Total amount of records in this collection: 1889624

de vries

Total number of results based on "de vries" and selected filters: 1130033

Select date field Start date

Broadcaster ✕ Genre (series) ✕ Keyword ✕ NEW +

voetbal (12010) muziek (11239) kinderen (9469) betogingen (8916) popmuziek (8333) tentoonstellingen (5291) politie (5139) toneel (5139) geschiedenis (5061) zangers (487) vluchtelingen (4605) poëzie (4541)

Wikidata Retrieval Service (?)

GreenLeft (Groenlinks)

Hafkamp, Hetty

Willems, Wilbert

Voortman, Linda

Dibi, Tofik

Gent, Ineke van

Broekhoven, Karel van

Barske, Anton

Lucas, Marjan

Brand, Arie van den

Select more

1 2 3 4 5 6 7 8 9 10 11 Next Last

DE NIEUWS BV; De nieuws BV (2017 - 2018) (23-01-2018)

Politiek, cultuur, opinie, humor en sport bij BNNVARA. Niet te missen elke werkdag van 12.00 uur tot 14.00 uur met Willemijn Veenhoven en Patrick Lodiers.

Fig. 9. Wikidata Retrieval Service - Select Persons

## Single Collection Search

Set collection (Audiovisual Collection) Set project (Test)

Searching in : Audiovisual Collection  
Total amount of records in this collection: 1889624

Wikidata Retrieval Service (?)

Hafkamp, Hetty; Dibi, Tofik; Barske, Anton

Search in: all fields

Total number of results based on Wikidata Retrieval Service selection: 15692

Fig. 10. Wikidata Retrieval Service - Search Query Selection