UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Using Linked Data to Track and Trace Processes of Canonization in early Modern Dutch Literature

Nijboer, H.; van Deinsen, L.; van Wissen, L.; Brouwer, J.; van Strien, T.; Blom, F.

Using Linked Data to Track and Trace Processes of Canonization in Early Modern Dutch Literature

Harm Nijboer[1], Lieke van Deinsen[2], Leon van Wissen[1], Judith Brouwer[3], Ton van Strien[4], and Frans Blom[1]

[1]University of Amsterdam
[2]KU Leuven
[3]Huygens ING
[4]Independent scholar

This article is about a pilot project in which we linked five web resources in order to get a firmer grip on processes of canonization in Dutch literature in the early eighteenth century. The project centers around the Panpoëticon Batavûm, an early eighteenth century portrait gallery of Dutch writers, initiated by the painter Arnoud van Halen (1673-1732) and continued by others. This hall of fame provides an early example of the historical canonization of Dutch writers. As a collection the Panpoëticon is no longer intact, but it has been digitally reconstructed as a website. We created a RDF representation of this website and combined these data with data derived from the DBNL, Ecartico, Onstage and the Short Title Catalogue Netherlands. Using these combined data sets we investigate whether the Panpoëticon reflects the popularity of playwrights in the Amsterdam theater and vice versa, but we also address potential geographical and sociocultural biases in the Panpoëticon. Besides generating new insights in processes of canonization in early modern Dutch literature, this article aims to provide a working example of the synergy that can be achieved by combining multiple data sets using Semantic Web technology.

1 Introduction

Fame. Few phenomena are so quintessential to our culture and yet, from an academic perspective, so precarious to deal with. Researchers dealing with fame and famous people are likely to run into what William Labov (1972) called an observer's paradox, since fame is basically created by calling it into being. Hence you can not observe people as being famous without contributing to their fame. This is particularly an issue in our own academic discipline, the history of literature and art, which has a long track record of establishing and re-establishing the fame of past writers and artists.

Figure 1: A selection of portraits that once belonged to the Panpoëticon, Rijksmuseum Amsterdam.

Along with popular views and partisanism, scholarship has been very instrumental in the shaping of the artistic and literary canons that, whether we like it or not, guide our views and understanding of art and literature in the past.

In the past decades literary and artistic canons have increasingly become subject to criticism and critical review. Aside from the objections made by (postmodern) critical theory that literary and artistic canons tend to reflect the views and interests of the dominant groups in culture and society, there is a growing awareness that such canons are likely to bias our understanding of past culture in its own context. And as a consequence the understanding of the dynamics behind processes of canonization is highly relevant to the history of art and literature (Gorak, 2013; Kolbas, 2001).

It will be clear that artistic and literary canons as social conventions should be distinguished – but not separated – from attempts to record and petrify such canons in writings, galleries and monuments. Such attempts have been numerous and can be traced back to Antiquity. In the realm of the arts the publication of Vasari's *Vite* (1550) is generally understood as a first attempt to define a canon of Italian Renaissance art. The example of Vasari was followed by Van Mander (1604) for painters from the Low Countries. Updates to Van Mander were provided by De Bie (1662) and Houbraken (1721). At the same time when Houbraken was using his writing skills to celebrate the painters, his fellow painter and friend Arnoud van Halen (1673-1732) decided to use his painting skills to celebrate the poets.

In the first decades of the eighteenth century Van Halen created a collector's cabinet that would eventually contain over 338 miniature portraits of Dutch authors (Figure 1). Throughout the eighteenth century this Panpoëticon Batavûm served as the *de facto* canon of Dutch literature (Leemans and Johannes, 2013; Van Deinsen, 2016). It was praised in hundreds of laudatory verses. And all these writings contributed to the authority of the Panpoëticon as a canon. After the death of Van Halen in 1732 the Panpoëticon was subsequently acquired by the broker and art lover Michiel de Roode in 1732, Arnoud de Jonghe in 1771, and the Leiden literary society *Kunst Wordt Door Arbeid Verkregen* [Art is Obtained by Labor] in 1772 (Figure 2). Under these new owners several old portraits were replaced and new likenesses were added to the cabinet.

Unfortunately, the cabinet was heavily damaged during the Leiden gunpowder disaster of 1807, although the individual portraits remained intact. However, the society, financially devastated by the disaster, was forced to put the collection of

Figure 2: The meeting hall of the literary society *Kunst Wordt Door Arbeid Verkregen* with in the middle, at the end of the table the cabinet of the Panpoëticon Batavum. Paulus Constantijn la Fargue, oil on canvas, 1774. (Rijksmuseum Amsterdam / Museum De Lakenhal Leiden)

paintings up for sale. In a final, desperate attempt to clear its debts, it offered the Panpoëticon to the newly founded Koninklijk Museum, a direct precursor of the Rijksmuseum Amsterdam, for the hefty sum of five thousand guilders. Despite the fact that nationalism was by then the thriving idea in politics and culture, interest in the eighteenth-century and rather encyclopedic collection was limited and the society's offer was firmly rejected twice. The collection was eventually sold in 1819 and again in 1849. The then owner, a profit-driven art broker, sold the portraits separately and they ended up in collections all across Europe (Van Deinsen, 2017, 2020). The wooden cabinet itself was lost.

Regret always comes later. Currently eighty-two of the original portraits have found their way to the Rijksmuseum in Amsterdam, others are part of different heritage collections or in the possession of private collectors. Of many the current whereabouts remain unknown. Meanwhile the Panpoëticon has been digitally 'reconstructed' as a website *Het Schrijverskabinet* (*www.schrijverskabinet.nl*) which gives a systematic overview of the portraits that were once in the cabinet as well as some additional information on the writers being portrayed. This website is of course an excellent tool to provide visitors with a feeling of what the cabinet was once about, but it falls short in providing researchers with structured data about the Panpoëticon and its inductees.

In this paper we will describe how we created a (linked) data representation of the Panpoëticon and we will explore its research potential. There were two major reasons why we thought that having a representation of the Panpoëticon in structured data would be useful. The first reason is that in addition to conventional historical research the quantitative and computational analysis of such material might be helpful in unveiling the less manifest forces, structures and biases behind processes of historical canonization as has been wittily illustrated by Skiena and Ward (2013). The second – but not less important – reason is that in an emerging landscape of linked data concerning art and literature in the Dutch Golden Age (Brouwer and Nijboer, 2018),

the addition of the Panpoëticon would be a valuable contribution *per se*. This collection is especially relevant since its genesis provides us with the unique opportunity to study the hybrid process of fame-making from the perspective of the cultural brokers involved. And as such it provides a little backdoor to escape – at least for a moment – the aforementioned observer's paradox.

2 The Panpoëticon in RDF

Considering the aforementioned objectives, the use of Semantic Web technology for creating a data representation of the Panpoëticon is the most obvious choice. Semantic Web technology consists of three main components: Unified Resource Identifiers (URIs), the Resource Description Framework (RDF), and the SPARQL query language. URIs (world wide unique identifiers) are used to represent things and concepts. Preferably URIs should take the form of an http(s) address, so that the things or concepts they represent, can be identified using the Web. RDF is a generic data model that describes data in subject-predicate-object triples, like Shakespeare (subject) has gender (predicate) male (object). In the RDF data model subject and predicate are always represented by a URI while the object may also be represented by a textual or a numerical value (literal). Hence, we can make the previous statement, using only URIs in a form like:

(1) *<http://viaf.org/viaf/96994048><http://schema.org/gender> <http://schema.org/Male>*.

For the sake of convenience we can shorten the lengthy 'http:etc' parts with commonly used prefixes, like:

(2) viaf:96994048 schema:gender schema:Male .

Since the object of a RDF triple can be the subject of another triple, the data model also allows for more complex statements like "Hamlet is a play by Shakespeare, who was born in 1564", which can be represented in RDF like:

(3) viaf:312343799 rdf:type schema:Play .
viaf:312343799 schema:name "Hamlet".
viaf:312343799 schema:author viaf:96994048 .
viaf:9699404 schema:name "William Shakespeare" .
viaf:9699404 schema:birthDate "1564" .

This feature of making complex statements by matching either the subject or the object of one triple to either the subject or object of another triple, makes the RDF data model very versatile, as data can be represented as a graph. And although certainly not all complexity can be easily accounted for in RDF, it has one other BIG advantage: ease of aggregation. Because every RDF dataset consists of triples, RDF datasets can easily be merged, even when they differ vastly in terms of content, terminology and scope. For this reason, the adaptation of Semantic Web technology has been particularly strong in fields where data is heterogeneous and semantically rich. The most obvious examples are general knowledge graphs like DBpedia and Wikidata. Another field that witnessed a steady adaptation of the RDF data model is cultural heritage. The collection databases of museums, libraries and archives typically cover data that is

diverse in structure and content. Meanwhile open data legislation increasingly urges publicly funded institutions to publish their data on the Web.

Large players in the Dutch cultural heritage field like the Rijksmuseum and the Royal Library have already published their collection databases in a RDF format. Obviously a RDF representation of the Panpoeticon would be a welcome addition to this emerging Semantic Web of Dutch cultural heritage. Especially since the Panpoëticon is already reconstructed as the *Schrijverskabinet* website and URIs should ideally point at a resource where the data about the entity being described, is served in both a human readable (HTML, i.e. a web page) and a machine readable (RDF) format. However, since this exercise was explicitly conducted as a pilot project, converting the *Schrijverskabinet* website into a full-fledged Semantic Web resource was not an option. Instead we used a less intrusive method by creating a separate data layer that represents the *Schrijverskabinet* in RDF. This layer serves both as a mere copy in terms of structure and relations between portrayed figures and articles about these persons, as well as an entry point that offers references to the *Schrijverskabinet* website and other datasets where possible.

Converting the *Schrijverskabinet* website into structured RDF data proved to be rather straightforward. Although the website does not provide structured data, it does use rather structured HTML, in the sense that individual data elements are recognizably ordered in the so called DOM tree of the HTML source code of the separate web pages, as is illustrated in Appendix 1 and Figure 3. For instance, the name of the portrayed person is taken from the content of the HTML h1 (heading) element, while an additional disambiguating description is taken from the h2 element. Similarly, information on the painter, the creation date of the portrait and its current location is taken from an array of labeled elements. Following this logic, the HTML was parsed into an intermediary data file of which a part is shown in Appendix 2.

Having a copy of the *Schrijverskabinet* website stored, the next step was to normalize and enrich the data, so that it can be converted to RDF. Part of this normalization is for instance the conversion of any dates that are given to a proper date description, so that the year range "1700 – 1732" can be expressed into a possible beginning and ending of respectively 1700-01-01 and 1732-12-31, which boosts 'queryablity' when working with the data. Also, if multiple painters worked on a portrait (e.g. if Arnoud van Halen first made a portrait, after which Jan Maurits Quinkhard augmented it), then this information is split so that we can attribute the portrait to two painters, each expressed with their own unique identifier (URI).

As mentioned above, these URIs define the resources in the world of RDF. Apart from properties that express for example the name or description of a resource, is there also a property that states the type of this resource, which points to a particular class from a specific ontology or vocabulary. This way, we can say that painter and collector Arnoud van Halen is referred to by a URI and is an instance of the type Person, which means that he is a member of everything in the world we classify as Person:

(4) *<https://data.create.humanities.uva.nl/id/schrijverskabinet/person/arnoud-van-halen>* rdf:type schema:Person .

Similarly, the (self) portrait of Arnoud van Halen is an instance of the type VisualArtwork:

(5) *<https://data.create.humanities.uva.nl/id/schrijverskabinet/artwork/arnoud-van-halen>* rdf:type schema:VisualArtwork .

Hugo de Groot

Rechtsgeleerde en schrijver

Delft 1583 – Rostock 1645

⬦⬦⬦

SCHILDER	**Arnoud van Halen**
DATERING	**1700 – 1720**
VINDPLAATS	**Rijksmuseum, Amsterdam**
ARTIKEL	**Geleerd multi-genie in een jungle van politieke spanningen** door Jan Waszink
DBNL-PROFIEL	**Hugo de Groot**

Aan Delft

Poogt Delft de Groot een graf te delven?
Hier is zijn lijk: maar voor het brein
Zijn alle graaven veel te klein:
Wie 't doen wil moet all d'aardt verwelven.
De geest is grooter dan de Groot.
De wijsheidt vult de werreldtkloot.

Jan Vos (1662), 'Op Hugo de Groot'. In: *Gedichten I*.

Figure 3: The portrait page of Hugo de Groot on the *Schrijverskabinet* website.
Source: *http://www.schrijverskabinet.nl/portret/hugo-de-groot/*

Instead of defining these concepts of Person and VisualArtwork ourselves, we make use of the classes from the well-known Schema.org vocabulary. Reasons to choose for Schema.org are its practical usability and the fact that this vocabulary is (getting) well-adopted amongst other providers of cultural heritage data, like the National Library of the Netherlands (KB), and the Netherlands Institute for Art History (RKD). In general, to advance the interoperability of RDF datasets it is good practice to reuse existing vocabularies, particularly for expressing the predicates in RDF triples and for the definition of classes (types) of the entities the dataset deals with.

The set of entities described by the *Schrijverskabinet* is rather limited and can be modeled into resources of the following classes: (1) schema:CreativeWork, for a reference to the original *Schrijverskabinet* portrait page, (2) schema:Person, for persons portrayed, painters, and contemporary authors that contributed to the *Schrijverskabinet*, (3) schema:Place, for birth and death places of persons portrayed, (4) schema:PublicationEvent, for the moment the portrait was added to the Panpoëticon, (5), schema:ScholarlyArticle, for contemporary articles listed on *www.schrijverskabinet.nl* about a particular portrayed person written by a present-day scholar, and (6) schema:VisualArtwork, for the portrait itself. Each of the modeled resources contains class specific properties to express additional information. The full set of RDF triples that is used to express data on for instance Hugo de Groot can be seen in Appendix 3.

The pipeline we used to map the *Schrijverskabinet* website into a RDF dataset is publicly available in a GitHub repository and can be rerun when the website is updated. The code and resulting dataset have also been published in the Zenodo research portal (Van Wissen and Nijboer, 2020).

3 Connecting resources: true Linked Data

Linked Data is simply about using the Web to create meaningful links between data from different sources (Bizer et al., 2011). RDF datasets are often considered Linked Data *per se*, since they use URIs to identify things, and link to external vocabularies to identify concepts. But that definition would be too confined. Linked Data becomes really interoperable when, and only when, entities described in a dataset are linked to entities in other datasets or to commonly used identifying URIs. It is about making explicitly clear that the John you are talking about is the same as the John someone else is talking about.

By convention the owl:sameAs predicate from the Web Ontology Language vocabulary is used to state that an entity in one dataset is the same as an entity in another dataset. And for this purpose it is highly recommended to refer to widely used identity providers like the Virtual Internet Authority File (VIAF). So when for instance my dataset states that:

(6) mydataset:John owl:sameAs viaf:40648922 .

And your dataset states that:

(7) yourdataset:Johannes owl:sameAs viaf:40648922 .

We can infer that:

(8) mydataset:John owl:sameAs yourdataset:Johannes .

Providing owl:sameAs links is a simple and effective way to identify entities across datasets. Yet, there is an ongoing discussion on whether the owl:sameAs predicate is the appropriate property for this task. In some circumstances this predicate is too blunt for the task, while in other circumstances its specific semantics may lead to undesired inferences (cf. Halpin et al., 2010; Idrissou et al., 2017). A full discussion of this issue goes well beyond the scope of the present paper and is in this context not very relevant, as potential problems usually arise at the fringes of the data landscape we want to explore. Furthermore, currently there is no viable alternative to the use of owl:sameAs links. They are still considered (part of) the 'glue that connects different datasets on the Semantic Web' (Beek et al., 2018).

The curators of the *Schrijverskabinet* already provided for each portrayed person a hyperlink to its corresponding page in the thesaurus of authors maintained by the Digital Library for Dutch Literature (DBNL), which is published by the National Library of the Netherlands (KB) as a separate dataset in their Linked Data Portal. When we started the project described in this paper, five inductees in the Panpoëticon did not have an entry in the DBNL thesaurus of authors yet. The DBNL, however, was so generous to create new entries for those persons. Today (almost) every person described at the *Schrijverskabinet* website is now linked to a corresponding page in the DBNL thesaurus of authors. These links to the DBNL make it possible to combine the RDF data derived from the *Schrijverskabinet* website with other RDF datasets. Contentwise the most relevant in this respect are the Short Title Catalogue Netherlands (STCN), ONSTAGE and ECARTICO.

The STCN is the Dutch retrospective bibliography for the period 1540-1800. The STCN contains bibliographic data on over 213 000 (and counting) titles that are still present in the collections of Dutch and several foreign (academic) libraries. This focus on physical books that survived the ravages of time always raises the question whether the STCN provides a representative overview of the output of the Dutch book industry prior to 1800. Recent research estimates that the STCN covers only 20% of the total production of printed works in the Dutch Republic (Pettegree and Der Weduwen, 2018, 2019). But in the case of literary works, with the possible exception of occasional poetry, this coverage should be estimated much higher since literary works have always been actively collected by libraries. Coverage in this field should be considered comprehensive enough for obtaining at least decent proxies for output and popularity. Authors of books in the STCN are linked to the Dutch Thesaurus for Author Names (NTA), which in turn is linked to VIAF. Both the STCN and the NTA have been made available by the KB as RDF datasets in their Linked Data Portal.

ONSTAGE, is a dataset containing the (almost) complete programming of the Amsterdam city theater from 1637 up until the present. The original programs have been enriched by links to structured representations of identified plays. The data allows researchers to trace the popularity of plays and playwrights over time and to track the rising popularity of Spanish drama in the seventeenth and French drama in the eighteenth century (Blom et al., 2020; Jautze et al., 2016). Persons (playwrights) and plays have been extensively linked to external identifiers. ONSTAGE is curated and hosted at the University of Amsterdam and is available as Linked Open Data.

ECARTICO is a prosopographical database with data on agents in the creative industries of the Low Countries from ca. 1475 until ca. 1725. The particular strength of the dataset is that it does not only deal with the who, what, when and where of the individual agents but that it also provides data on how these agents were related, either by kinship or other social relations. Although ECARTICO's main objective
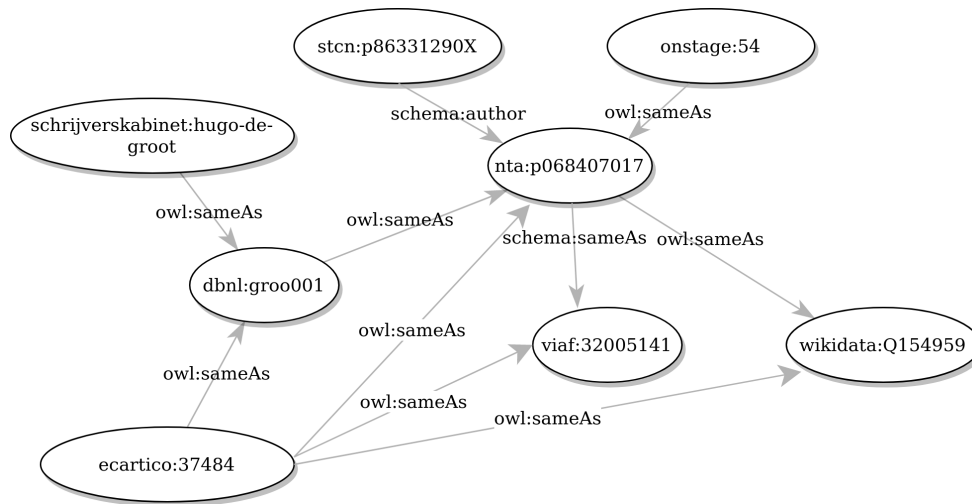
Figure 4: Schema of the interlinkage of the *Schrijverskabinet* to other resources.

is to establish a prosopography (a systematic overview of people, their attributes and relations) it does also provide extensive linkage to external resources like VIAF, the NTA, the DBNL and others (Brouwer and Nijboer, 2018). To support the current project the editors of ECARTICO included all inductees of the Panpoëticon. ECARTICO is curated and hosted at the University of Amsterdam in cooperation with Huygens ING and is available as Linked Open Data.

Figure 4 shows how the *Schrijverskabinet* is linked to the resources mentioned above by only providing a link to the DBNL authors thesaurus. The most obvious is the linkage between ECARTICO and the *Schrijverskabinet*. Although these datasets do not directly link to each other, person entities described in both datasets do share a link to the DBNL authors thesaurus. As an intermediary both the DBNL and ECARTICO link the *Schrijverskabinet* to the NTA. This might seem redundant, but ECARTICO might provide identifications that so far have escaped the attention of the curators of the DBNL and the other way around. The NTA in turn is linked to the STCN and ONSTAGE. At the 'end' of the chain of owl:sameAs links appears Wikidata, the global database that anyone can contribute to. The author's resource on Wikidata might serve in this context as a global low-level entry point for third parties to store, connect and publish data.

4 Some results

The benefit of linking the *Schrijverskabinet* to a web of resources is that one can harvest information about a person from various datasets that all have their specific expertise and scope. Biographical data is for instance better documented and structured in ECARTICO, while bibliographic information can be pulled out of the STCN. Finally, ONSTAGE can be used to check whether the most popular playwrights of the seventeenth century managed to obtain a canonical status by becoming inductees in the Panpoëticon.

Below we will present three use cases for which we queried the *Schrijverskabinet* RDF data in combination with other datasets. To make RDF data queryable it has to be stored in a triple store which can be accessed over the Web through a SPARQL

Figure 5: Birthplaces of inductees in the Panpoëticon.

endpoint. An endpoint for the NTA, STCN and DBNL datasets is provided by the KB. For the other datasets a (temporary) endpoint is provided by the University of Amsterdam, while a more permanent endpoint will be provided by the Golden Agents research infrastructure in the near future. SPARQL is the dedicated language to query RDF data. The logic behind SPARQL and its syntax are well documented, both in print and on the Web. But in practice, it is for the greater part learned by doing, as with most computer languages. So, for learning purposes as well as for the sake of reproducibility and replicability, the SPARQL queries that generated these outcomes used in the cases below, are included in the appendices.

4.1 Mapping the birthplaces of inductees in the Panpoëticon

To address a potential geographical bias in the Panpoëticon we would like to know more about the places of birth of the inductees. Since structured data on birthplaces is not provided by the original *Schrijverskabinet* data, we took the birthplaces of the inductees from ECARTICO. The query we used for this purpose is included in Appendix 4, while the result is mapped in Figure 5.

As was expected the Panpoëticon seems pretty biased towards writers born in the western part of the Dutch Republic. Closer examination of the figures reveal that especially Amsterdam (77 inductees) was well represented. The second city on the list, Dordrecht, gave birth to only 17 literary heroes. People born in the eastern parts of the Netherlands (Drenthe, Twente, Achterhoek) were conspicuously absent in the Panpoëticon.

4.2 Inductees and their published works

The next thing we wanted to know is how well the inductees of the Panpoëticon are represented in the STCN. We designed a query (Appendix 5) that asks for the number of titles in the STCN for each writer in the Panpoëticon. The query runs over the *Schrijverskabinet* and the STCN using ECARTICO as an intermediary to obtain the applicable NTA URIs. The query returned 272 results, with a top ten of prolific writers that one would more or less expect (Vondel, Erasmus, Grotius, etc.) But more

Table 1: Most popular playwrights in the Amsterdam City Theater checked against the inductees in the Panpoëticon.

Number of stagings	Name	Inductee
973	Joost van den Vondel	Yes
809	Isaak Vos	No
697	Pieter Bernagie	No
576	Dirck Buysero	Yes
544	Pieter de la Croix	No
488	Thomas Asselijn	No
454	Frans Rijk	Yes
453	Jacob van Rijndorp	No
449	Jan Vos	Yes
349	Adriaan Bastiaensz. de Leeuw	No
346	Dirck Pietersz. Heynck	No
336	Willem van der Hoeven	Yes
333	Andries Pels	Yes
304	Reinier Bontius	No
301	Enoch Krook	No
283	Daniel Kroon	No
279	Pieter Langendijk	Yes
272	Jan Pluimer	No
255	Lodewijk Meyer	No
249	David Lingelbach	Yes
244	Pieter Cornelisz Hooft	Yes
217	Joan Dullaart	No
217	Willem Godschalk van Focquenbroch	Yes
201	Claas Bruin	Yes
191	Ysbrand Vincent	Yes

surprising is that of 124 writers that were inducted in the Panpoëticon no works can be found in the STCN. This can partly be explained by the fact that many poets had their poems only published in anthologies. Apparently, not having published single authored works was not an obstruction for writers to gain canonical status in the early modern period.

4.3 Canonization of popular playwrights

Finally we wanted to investigate whether the most popular playwrights of the seventeenth and eighteenth centuries were also inductees of the Panpoëticon. For this we counted for each playwright the number of stagings of their plays in the Amsterdam City Theater prior to 1785 which might serve as a proxy for popularity. In addition we checked for each playwright whether they were inducted in the Panpoëticon or not. This whole operation can be executed by one single SPARQL query (Appendix 6) running over the *Schrijverskabinet*, ONSTAGE and ECARTICO as an intermediary. The resulting Top 25 is displayed in Table 1.

The results of this exercise are at least remarkable, since it shows that being a successful playwright was no guarantee for reaching a canonical status. An explanation might be that many of the popular playwrights that failed to get inducted in the

Panpoëticon, were mainly active as translators of French and Spanish drama. But in that case, one can also state that the Panpoëticon failed to pay tribute to a group of writers that had an enormous impact on Dutch literary culture in the seventeenth and eighteenth centuries (cf. Jautze et al., 2016).

5 Final remarks

In this paper we described a pilot project in which we converted the *Schrijverskabinet* website, a digital reconstruction of the Panpoëticon Batavûm, into a RDF dataset. We were able to combine this dataset with other RDF datasets using the links to the DBNL authors thesaurus provided by the *Schrijverskabinet*. By querying the combined datasets we were able to generate contextual statistics for the Panpoëticon that might be helpful in generating new insights in processes of canonization in early modern Dutch literature. As such the project provides a working example of the synergy that can be achieved by combining multiple datasets using Semantic Web technology.

All three use cases we presented in this study, showed that the Panpoëticon was in no way an unbiased canon of Dutch literature. Notwithstanding the fact it was presented and perceived as an undisputed monument for Dutch literature by many eighteenth-century contemporaries (Van Deinsen, 2017), our detailed analysis of its content shows the collection was, in fact, heavily focused on authors from Amsterdam, included several overrated poets with a meager literary publication track and underrated successful dramatists that were seen as mere translators. The process of literary canonization, as materialized in the Panpoëticon, thus, was by no means only based on the literary quality of the inductees. All of this was of course more or less known by the small circle of initiates of the Panpoëticon, but data analysis reveals these biases more explicitly and to a point beyond doubt. As such, our pilot project has highlighted the added value of combining resources in charting the complex and oftentimes hidden mechanisms behind literary canon formation. This paves the way for a similar approach of other canonization projects, both in the past and present.

With regard to the specific case of the Panpoëticon: if the cabinet did not contain an unbiased canon of Dutch literature, it becomes all the more interesting to reveal hitherto unidentified factors contributing to the collection's formation. Future research in, for example, the social status, social aspirations and networks of its initiator, its subsequent owners and maintainers and other stakeholders involved could provide new insights in processes of early modern canon formation. Data for such large scale prosopographical analysis of stakeholders and inductees is to a large part already available or is increasingly becoming available. The Linked Data framework presented in this paper ensures that data on the Panpoëticon can be used in conjunction. On a more technical level embedding this framework in the original context of the Schrijverskabinet website is an option that seriously should be considered.

Finally, we want to stress that the success of this project owes a lot to the extensive communication and collaboration between data providers and researchers. To align the data there has been an extensive and animated email exchange between the editors of the *Schrijverskabinet* and the editors of ECARTICO. Making the *Schrijverskabinet* successfully interlinked by pointing at only one domain, also owes much to the generosity of the DBNL to create new identifiers in their thesaurus for those inductees, who were not represented there yet. Last but not least, we have to mention the workshop at the Amsterdam Time Machine / Golden Agents Datasprint (March 12th, 2020) where we had lively discussions about the research potential of this data experiment. All of this

exemplifies once again that using Linked Data in humanities research is only partly a technological challenge. In the end, it is about connecting scholars and scholarship.

Acknowledgements

References

Beek, W., J. Raad, J. Wielemaker, and F. Van Harmelen
  2018. sameas.cc: The closure of 500m owl:sameas statements. In *The Semantic Web. ESWC 2018. Lecture Notes in Computer Science*, A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, eds., volume 10843, Pp. 65–80. Cham: Springer.

Bizer, C., T. Heath, and T. Berners-Lee
  2011. Linked data: The story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, A. Sheth, ed., Pp. 205–277. Hershey, PA: IGI Global.

Blom, F., H. Nijboer, and R. Van der Zalm
  2020. Onstage, the online data system of theatre in amsterdam from the golden age to today. *Research Data Journal for the Humanities and Social Sciences*, 5(2):27–40.

Brouwer, J. and H. Nijboer
  2018. Golden agents. a web of linked biographical data for the dutch golden age. In *Proceedings of the Second Conference on Biographical Data in a Digital World 2017, Linz, Austria, November 6-7, 2017.*, A. Antske Fokkens, S. Ter Braake, R. Sluijter, P. L. Arthur, and E. Wandl-Vogt, eds., CEUR Workshop Proceedings, Pp. 33–38. CEUR-WS.org.

De Bie, C.
  1662. *Het gulden cabinet van de edel vry schilderconst*. Antwerpen: Jan Meyssens & Juliaen van Montfort.

Gorak, J.
  2013. *The making of the modern canon: genesis and crisis of a literary idea*. London: Bloomsbury Academic.

Halpin, H., P. J. Hayes, J. P. McCusker, D. L. McGuinness, and H. S. Thompson
  2010. When owl:sameas isn't the same: An analysis of identity in linked data. In *The Semantic Web – ISWC 2010*, P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, and I. Horrocks, eds., volume 6496 of *Lecture Notes in Computer Science*, Pp. 305–320. Berlin & Heidelberg: Springer.

Houbraken, A.
  1718–1721. *De groote schouburgh der Nederlantsche konstschilders en schilderessen*. 's Gravenhage: J. Swart, C. Boucquet & M. Gaillard.

Idrissou, A., R. Hoekstra, F. Van Harmelen, A. Khalili, and P. Van Den Besselaar
  2017. Is my:sameas the same as your:sameas? lenticular lenses for context-specific identity. In *Proceedings of the Knowledge Capture Conference, K-CAP 2017*, Pp. 1–8, New York. Association for Computing Machinery, Inc.

Jautze, K., L. Álvarez Francés, and F. Blom
  2016. Spaans theater in de Amsterdamse Schouwburg (1638-1672). Kwantitatieve en kwalitatieve analyse van de creatieve industrie van het vertalen. *De Zeventiende Eeuw. Cultuur in de Nederlanden in interdisciplinair perspectief*, 32(1):12–39.

Kolbas, E.
  2001. *Critical Theory and the Literary Canon*. Boulder: Westview Press.

Labov, W.
  1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Leemans, I. and G. Johannes
  2013. *Worm en donder. Geschiedenis van de Nederlandse literatuur 1700-1800: De Republiek*. Amsterdam: Bert Bakker.

Pettegree, A. and A. Der Weduwen
  2018. What was published in the seventeenth-century dutch republic? In *Livre. Revue historique*, Pp. 1–22. Paris: Société bibliographique de France.

Pettegree, A. and A. Der Weduwen
  2019. *The Bookshop of the World: Making and Trading Books in the Dutch Golden Age*. New Haven and London: Yale University Press.

Skiena, S. and C. Ward
  2013. *Who's Bigger? Where Historical Figures Really Rank*. New York: Cambridge University Press.

Van Deinsen, L.
  2016. *The Panpoëticon Batavûm. The Portrait of the Author as a Celebrity*. Amsterdam: Rijksmuseum.

Van Deinsen, L.
  2017. *Literaire erflaters: Canonvorming in tijden van culturele crisis, 1700-1750*. Hilversum: Verloren.

Van Deinsen, L.
  2020. A tale of images and words: The panpoëticon batavûm and the failed attempt to publish a history of dutch literature. *The Rijksmuseum Bulletin*, 68(1):54–63.

Van Mander, C.
  1604. *Het Schilder-Boeck*. Haerlem: Paschier van Wesbusch.

Van Wissen, L. and H. Nijboer
  2020. Schrijverskabinet rdf (version 1.0). Dataset.

Vasari, G.
  1550. *Le vite de piu eccellenti architetti, pittori, et scultori Italiani, da Cimabue insino a' tempi nostri*. Firenze: Lorenzo Torrentino.

Code and Data Availability

The pipeline used to convert the *Schrijverskabinet* into RDF, as well as the data itself, is available at:
https://github.com/LvanWissen/schrijverskabinet-rdf

Version 1.0 of the *Schrijverskabinet* RDF dataset has been published in the Zenodo research portal as Van Wissen and Nijboer (2020) and is available at:
https://doi.org/10.5281/zenodo.3835559

RDF versions of the DBNL authors thesaurus, the NTA and the STCN are provided by the KB in their Linked Data Portal:
http://data.bibliotheken.nl

ONSTAGE is available in HTML/RDFa at:
http://www.vondel.humanities.uva.nl/onstage/

ECARTICO is available in HTML/RDFa at:
http://www.vondel.humanities.uva.nl/ecartico/

Most of the aforementioned RDF datasets are aggregated (and regularly updated) in the Golden Agents research infrastructure at:
https://data.goldenagents.org/

Please note that many of the resources used in this study are actively maintained and regularly updated. Hence, rerunning experiments will probably yield slightly different results.

Appendices

Appendix 1: Fragment of the (slightly tidied) HTML source code of the web page describing the portrait of Hugo de Groot (Grotius) at the *Schrijverskabinet* website

```html
<div id="portrait-info">
  <h1>Hugo de Groot</h1>
  <h2>Rechtsgeleerde en schrijver</h2>
  <div class="date-of-birth-and-death">
    Delft 1583 - Rostock 1645
  </div>
  <div class="details">
    <div class="label">Schilder</div>
    <div class="data">Arnoud van Halen</div>
    <div class="label">Datering</div>
    <div class="data">1700 - 1720 </div>
    <div class="label">Vindplaats</div>
    <div class="data">
      <a href="http://hdl.handle.net/10934/RM0001.COLLECT.512936">
      Rijksmuseum, Amsterdam</a>
    </div>
    <div class="label">Artikel</div>
    <div class="data"><a
      href="http://www.schrijverskabinet.nl/artikel/hugo-de-groot/">
      Geleerd multi-genie in een jungle van politieke spanningen
      <span>door Jan Waszink</span></a>
    </div>
    <div class="label">DBNL-profiel</div>
    <div class="data"><a
      href="http://www.dbnl.org/auteurs/auteur.php?id=groo001">
      Hugo de Groot</a>
    </div>
  </div>
</div>
```

Appendix 2: Data coming from the *Schrijverskabinet* website for the portrait of Hugo de Groot stored as JSON

```
 1  "http://www.schrijverskabinet.nl/portret/hugo-de-groot/": {
 2      "title": "Hugo de Groot",
 3      "subtitle": "Rechtsgeleerde en schrijver",
 4      "bio": "Delft 1583 \u2013 Rostock 1645",
 5      "painter": "Arnoud van Halen",
 6      "date": "1700 \u2013 1720",
 7      "origin": {
 8          "name": "Rijksmuseum, Amsterdam",
 9          "url": "http://hdl.handle.net/10934/RM0001.COLLECT.512936"
10      },
11      "article": {
12          "name": "Geleerd multi-genie ...",
13          "url": "http://www.schrijverskabinet.nl/artikel/hugo-de-groot/"
14      },
15      "dbnl": {
16          "name": "Hugo de Groot",
17          "url": "http://www.dbnl.org/auteurs/auteur.php?id=groo001"
18      },
19      "quote": "Aan Delft ...",
20      "depiction": "http://www.schrijverskabinet.nl/[...]/Hugo-de-Groot-
        Jong-1.jpg",
21      "artdepiction": "http://www.schrijverskabinet.nl/[...]/Hugo-de-Groot
        -Jong.jpg"
22  }
```

Appendix 3: RDF Turtle representation of the webpage describing Hugo de Groot in the *Schrijverskabinet*

```
1  @base <https://data.create.humanities.uva.nl/id/schrijverskabinet/>.
2  @prefix skb: <http://www.schrijverskabinet.nl/>.
3  @prefix skbim: <http://www.schrijverskabinet.nl/wp-content/uploads/>.
4  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
5  @prefix schema: <http://schema.org/> .
6  @prefix owl: <http://www.w3.org/2002/07/owl#> .
7  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
8
9  <person/hugo-de-groot>
10     a schema:Person ;
11     schema:birthDate "1583"^^xsd:gYear ;
12     schema:birthPlace _:delft ;
13     schema:deathDate "1645"^^xsd:gYear ;
14     schema:deathPlace _:rostock ;
15     schema:disambiguatingDescription "Rechtsgeleerde en schrijver" ;
16     schema:mainEntityOfPage <skb:portret/hugo-de-groot/> ;
17     schema:name "Hugo de Groot" ;
18     schema:subjectOf <skb:artikel/hugo-de-groot/>,
19                      <artwork/hugo-de-groot> ;
20     owl:sameAs <http://data.bibliotheken.nl/id/dbnla/groo001> ;
21     foaf:depiction <skbim:2016/03/Hugo-de-Groot-Jong-1.jpg> .
22
23  <skb:portret/hugo-de-groot/> a schema:CreativeWork ;
24     schema:mainEntity <person/hugo-de-groot> ;
25     schema:text """Aan Delft ...""" .
26
27  <skb:artikel/hugo-de-groot/>
28      a schema:ScholarlyArticle ;
29      schema:about <person/hugo-de-groot> ;
30      schema:author <person/jan-waszink> ;
31      schema:name "Geleerd multi-genie in een jungle van politieke
       spanningen" .
32
33  # NOTE: The skb: and skbim: prefixes are used to make the code more
        legible. They may not work in real world applications.
```

Appendix 4: SPARQL query to get the birthplaces of the inductees in the Panpoëticon

```
1  # The query is executed at: https://data.create.humanities.uva.nl/
       sparql
2  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3  PREFIX sem: <http://semanticweb.cs.vu.nl/2009/11/sem/>
4  PREFIX owl: <http://www.w3.org/2002/07/owl#>
5  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
6  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
7  PREFIX schema: <http://schema.org/>
8  SELECT
9  COUNT(DISTINCT(?ecarticoPerson)) AS ?num
10 ((xsd:decimal(COUNT(DISTINCT(?ecarticoPerson))))/10+5) AS ?radius
11 ?birthPlaceName
12 ?lat
13 ?long
14 WHERE {
15
16   # First, fetch all persons and their portraits from the panpoeticon
17   GRAPH <https://data.create.humanities.uva.nl/id/schrijverskabinet/> {
18      ?person a schema:Person;
19             owl:sameAs ?dbnlPerson ;
20             schema:subjectOf ?portrait .
21   ?portrait a schema:VisualArtwork .
22   }
23
24   # We use the ECARTICO dataset to fetch the birthPlaces
25   GRAPH <https://data.create.humanities.uva.nl/id/ecartico/> {
26
27   ?ecarticoPerson a schema:Person ;
28           owl:sameAs ?dbnlPerson .
29
30   ?ecarticoPerson schema:birthPlace ?birthPlace .
31
32   ?birthPlace schema:name ?birthPlaceName ;
33             schema:geo [ schema:latitude ?lat ;
34                          schema:longitude ?long ] .
35   }
36
37
38 } GROUP BY ?birthplace ?birthPlaceName ?lat ?long
39 ORDER BY DESC(?num) ASC(?birthPlaceName)
```

Appendix 5: SPARQL query to list the inductees in the Panpoëticon and the number of their works catalogued by the STCN

```
1  # The query is executed at: http://data.bibliotheken.nl/sparql
2  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
3  PREFIX void: <http://rdfs.org/ns/void#>
4  PREFIX owl: <http://www.w3.org/2002/07/owl#>
5  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
6  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
7  PREFIX schema: <http://schema.org/>
8  PREFIX sem: <http://semanticweb.cs.vu.nl/2009/11/sem/>
9  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
10
11 SELECT DISTINCT ?worksInSTCN ?name ?authorNTA WHERE {
12
13   SERVICE <https://data.create.humanities.uva.nl/sparql> {
14   SELECT ?name ?authorNTA WHERE {
15     GRAPH <https://data.create.humanities.uva.nl/id/schrijverskabinet/>
        {
16       ?person a schema:Person;
17               schema:name ?name ;
18               schema:subjectOf [ a schema:VisualArtwork ] ;
19               owl:sameAs ?authorDBNL .
20     }
21
22     # We use the ECARTICO dataset to jump to the NTA
23     GRAPH <https://data.create.humanities.uva.nl/id/ecartico/> {
24       ?ecarticoPerson a schema:Person ;
25               owl:sameAs ?authorDBNL, ?authorNTA .
26
27       FILTER(CONTAINS(STR(?authorNTA), 'http://data.bibliotheken.nl/id/
      thes'))
28
29     }
30   }
31   }
32
33   {
34   SELECT (COUNT(?work) AS ?worksInSTCN) ?authorNTA WHERE {
35
36       ?authorNTA a schema:Person .
37
38       ?work foaf:isPrimaryTopicOf [ a foaf:Document ;
39                                     void:inDataset <http://data.
      bibliotheken.nl/id/dataset/stcn> ] .
40
41       {
42         ?work schema:author/schema:author ?authorNTA .
43       } UNION {
44         ?work schema:contributor/schema:contributor ?authorNTA .
45       }
46
47   } GROUP BY ?authorNTA
48
49
50   }
51
52 } ORDER BY DESC(?worksInSTCN)
```

58

Appendix 6: SPARQL query to list the most popular playwrights prior to 1785 (measured by the number of stagings of their plays in the Amsterdam City theater) and their presence in the Panpoëticon

```
1  # The query is executed at: https://data.create.humanities.uva.nl/
       sparql
2  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3  PREFIX owl: <http://www.w3.org/2002/07/owl#>
4  PREFIX schema: <http://schema.org/>
5  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
6  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
7
8  SELECT ?count ?name ?inductee WHERE {
9    {
10   SELECT (COUNT(DISTINCT ?show) AS ?count) ?name ?personONSTAGE WHERE {
11
12     # Ecartico is used to chain the Schrijverskabinet and OnStage
       together
13     GRAPH <https://data.create.humanities.uva.nl/id/ecartico/> {
14         ?personECARTICO a schema:Person ;
15                         schema:name ?name ;
16                         owl:sameAs ?externalURI, ?personDBNL .
17     }
18
19     GRAPH <https://data.create.humanities.uva.nl/id/onstage/> {
20         ?personONSTAGE a schema:Person ;
21                    owl:sameAs ?externalURI . # this links them
       together
22
23         ?playURI schema:creator ?personONSTAGE . # a play is written by
        an author
24
25         ?show schema:subEvent/schema:workPerformed ?playURI ; # a play
       is performed in a show
26                  schema:startDate ?date . # on a date
27
28         FILTER (?date <= "1785"^^xsd:gYear)  # filter to exclude
       contemporary performances
29     }
30
31   } GROUP BY ?name ?inductee ?personONSTAGE ORDER BY DESC(?count)
32 }
33   # If person is also included in the Panpoëticon, set 'inductee' to '
       Yes'
34   BIND(IF(EXISTS{
35     GRAPH <https://data.create.humanities.uva.nl/id/schrijverskabinet/>
        {
36         ?person a schema:Person ;
37                 schema:subjectOf [ a schema:VisualArtwork ] . }
38         ?person owl:sameAs/^owl:sameAs ?personONSTAGE .
39   }, "Yes", "No") AS ?inductee)
40
41 } LIMIT 25
```