

What we learned from preparing for a Linked Open Data hackathon

Experiences from a first timer team of the KB, national library of the Netherlands

*Olaf Janssen – olaf.janssen@kb.nl – @ookgezellig – 0000-0002-9058-9941,
Mirjam Cuper - Mirjam.Cuper@kb.nl – @CuperMirjam – 0000-0003-0187-9873,
Thomas Haighton –Thomas.Haighton@KB.nl – @ThomasHaighton,
Martijn Pluim – Martijn.Pluim@KB.nl,
Roosmarijn de Groot – Roosmarijn.deGroot@KB.nl – @RoosmarijndeG,
Daniëlle Jansen - Danielle.Jansen@KB.nl – @ecritures,
Koninklijke Bibliotheek, national library of the Netherlands
Lotte van den Eertwegh - lotte.van.den.eertwegh@outlook.com, Dutch Ministry of Internal Affairs*

The HackaLOD¹ is an annual one day hackathon of the heritage field in the Netherlands focusing on linked open data (LOD). Teams of programmers, historians, designers, and other heritage and data experts are challenged to create new, creative applications with collection data from heritage institutions, resulting in new insights into Dutch heritage by connecting collections.

For the 2020 edition we formed a KB² team to participate for the first time, bringing together colleagues from various departments, whose knowledge about SPARQL³, Wikidata⁴ and other LOD technologies ranged from basic to intermediate. As we never participated in a hackathon before and we are no LOD gurus who could hack together an application within a single day, we needed to make the necessary preparations well in advance. Due to COVID19 regulations the 2020 HackaLOD was repeatedly postponed, now to November 2022, but that did not discourage us from continuing our preparations, as team based learning from these preparations was a goal in itself.

First, we discussed the boundary conditions for our project. With the KB collection as our natural starting point, a requirement for the HackaLOD was to reuse LOD from other institutions as well. Furthermore, we wanted to link the past to the present, bringing heritage into a 21st century context. Finally, inspired by the success of KB's *Medieval Meme Generator*⁵, we wanted to build a fun and not-too-serious application. We aimed for collection objects that allowed us to build our own rich LOD cloud, not only by interlinking entities (people, locations, datetimes, events etc.) within our own datasets, but especially by connecting to entities in LOD endpoints of other parties as well.

We found the *Album amicorum of Jacob Heyblocq*⁶ [abbrev. *AAJH*]⁶ to be an excellent starting point. This 17th century book of friendship⁷ was created by Jacob Heyblocq (1623-1690), poet and headmaster of the Amsterdam Grammar School⁸. It holds contributions by Rembrandt⁹, Vondel¹⁰, Huygens¹¹, Cats¹² and 137 other persons from the Dutch Golden Age¹³. This allowed us to make semantic links between Heyblocq and all album contributors. As many of these contributors are described in LOD databases of other parties, it allowed us to retrieve additional information from those sources. Furthermore, most contributions are signed with both a placename (city) and a date, allowing us to make connections between persons, locations and dates. Finally, as some contributors maintained alba themselves, it should theoretically be possible to make connections to those alba and their contributors, contributions, locations and dates, adding even more nodes to our AAJH-LOD-cloud.

To build this cloud, we used Wikidata as our storage and created Wikidata items for the album itself¹⁴, for all contributions¹⁵ and for all contributors¹⁶. We did this by manual input via the Wikidata web interface and by semi-automated input via OpenRefine¹⁷, adding batches of data from local Excel sheets. We made sure all the entities were properly interconnected. This way, we were able to request data directly linked to the contributors, such as their genders, places and dates of birth and

death, portraits, parents, spouses, children, occupations and working locations. Obviously, this kind of data is much more complete for famous contributors such as Rembrandt or Huygens¹⁸ than for more obscure contributors such as P. Waterpas or J. van Bonevaal¹⁹.

From Wikidata we were also able to retrieve other artworks (mostly paintings and etchings) made by the contributors²⁰. Next, we manually added external identifiers to as many of our Wikidata items as possible²¹, at the same time relying upon both human and non-human (ie. bot) Wikidata community members to contribute to this effort as well. This allowed us to retrieve data from external databases (such as DBNL²² and eCartico²³,) via federated SPARQL queries. This way we were eg. able to find literary works written by the contributors or even their marriage or death certificates. Finally, we SPARQLed a number of LOD endpoints directly (including Europeana²⁴), without using Wikidata, allowing us to add even more data to our LOD cloud.

As we aimed to link the past to the present in a fun manner, we wanted to manifest our AAJH-LOD-cloud via a contemporary, lightweight frontend. Because the Heyblocq album can be considered a 'Facebook avant-la-lettre', we choose to partly recreate the user interface and interaction of modern popular social media, but now fully running on historical data from our cloud. Working title: *Forum Amicorum*²⁵, currently an early prototype.

To inspire digital scholars to play around with cultural heritage LOD in general and SPARQL in particular, and to encourage them to participate in hackathons, we would like to share four key insights on what we learned from building our AAJH-LOD-cloud and the *Forum Amicorum* prototype:

1) *Collectively working on a targeted, real-life project is an excellent way to improve your SPARQL skills.* Prior to the project, some team members found SPARQL rather intimidating and had limited knowledge, but thanks to explanations and query examples of more experienced members, their skills improved rapidly. Group based practical learning works better than individual theoretical learning.

2) *Playing with query examples beats reading documentation.* During our preparations we came across unfamiliar SPARQL endpoints and needed to find out how they worked. In some cases, the available documentation was of poor quality, ranging from non-existent to way too elaborate. Often informal playing with query examples rather than reading the formal manual proved a more efficient way for understanding the new SPARQL source. A positive example is Wikidata, with lots of queries being publicly shared²⁶.

3) *Real life contact with database experts is even better.* During our project we discovered eCartico²³, a new source (for us) that seemed to hold lots of useful information for our LOD cloud. Rather than trying to figure it out ourselves, we reached out to an expert of this database. During a very pleasant online meeting he explained us the ins and outs and answered all our questions, which saved us many hours.

4) *Sometimes SPARQL is not enough.* Despite the growing number of SPARQL enabled databases, there are still many that can only be accessed via their REST APIs, or even only by web scraping or other less advanced methods. Therefore, having a wide range of data gathering and processing skills is very useful for participating in a LOD hackathon.

*This article is made available under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license.*²⁷

-
- ¹ HackalOD, <https://hackalod.com/>
- ² KB, national library of the Netherlands, <https://www.kb.nl>
- ³ SPARQL, <https://www.w3.org/TR/sparql11-query/>
- ⁴ Wikidata, <https://www.wikidata.org>
- ⁵ KB Medieval Meme Generator, <https://www.medievalmemes.org/>
- ⁶ AAJH, Album amicorum of Jacob Heyblocq, <https://www.kb.nl/heyblocq>
- ⁷ Album amicorum, https://hob.gseis.ucla.edu/Resources/Knipprath_Reference.html and <https://www.kettererkunst.com/dict/album-amicorum-stambuch-memory-book-or-friendship-book.php>
- ⁸ Jacob Heyblocq, https://nl.wikipedia.org/wiki/Jacobus_Heiblocq
- ⁹ Rembrandt van Rijn, <https://en.wikipedia.org/wiki/Rembrandt>
- ¹⁰ Joost van den Vondel, https://en.wikipedia.org/wiki/Joost_van_den_Vondel
- ¹¹ Constantijn Huygens, https://en.wikipedia.org/wiki/Constantijn_Huygens
- ¹² Jacob Cats, https://en.wikipedia.org/wiki/Jacob_Cats
- ¹³ The Dutch Golden Age, https://en.wikipedia.org/wiki/Dutch_Golden_Age
- ¹⁴ Wikidata item of the album amicorum of Jacob Heyblocq, <https://www.wikidata.org/wiki/Q72752496>
- ¹⁵ List of contributions in the album amicorum of Jacob Heyblocq, Wikidata, <https://w.wiki/4p3U>
- ¹⁶ List of contributors to the album amicorum of Jacob Heyblocq, Wikidata, [https://w.wiki/F4\\$](https://w.wiki/F4$)
- ¹⁷ OpenRefine and Wikidata, <https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine>
- ¹⁸ Richly filled Wikidata items for Rembrandt <https://www.wikidata.org/wiki/Q5598> and for Huygens <https://www.wikidata.org/wiki/Q560746>
- ¹⁹ Poorly filled Wikidata items for P. Waterpas <https://www.wikidata.org/wiki/Q80678239> or Joannes van Bonevaal <https://www.wikidata.org/wiki/Q80695895>
- ²⁰ SPARQL queries in Wikidata for contributors of the Heyblocq album, <https://github.com/KBNLresearch/Hackalod/blob/main/queries/wikidata-AAJH-queries.md>
- ²¹ Wikidata external identifiers related to contributors of the Heyblocq album: ISNI ID, <https://www.wikidata.org/wiki/Property:P213>; VIAF ID, <https://www.wikidata.org/wiki/Property:P214>; Dutch National Thesaurus for Authors ID, <https://www.wikidata.org/wiki/Property:P1006>; Europeana Entity, <https://www.wikidata.org/wiki/Property:P7704>; DBNL author ID, <https://www.wikidata.org/wiki/Property:P723>; Biografisch Portaal number, <https://www.wikidata.org/wiki/Property:P651>; RKDartists ID, <https://www.wikidata.org/wiki/Property:P650>; Ecartico ID, <https://www.wikidata.org/wiki/Property:P2915>; Library of Congress authority ID, <https://www.wikidata.org/wiki/Property:P244>
- ²² DBNL title database, <http://data.bibliotheken.nl/doc/dataset/dbnl>
- ²³ eCartico, a comprehensive collection of structured biographical data concerning painters, engravers, printers, book sellers, gold- and silversmiths and others involved in the 'cultural industries' of the Low Countries in the sixteenth and seventeenth centuries, <https://data.create.humanities.uva.nl/page/ecartico/> and <http://www.vondel.humanities.uva.nl/ecartico>
- ²⁴ Europeana SPARQL endpoint, <http://sparql.europeana.eu/> and <https://pro.europeana.eu/page/sparql>
- ²⁵ Development prototype and early mockups of *Forum Amicorum*: http://forum_amicorum.kbresearch.nl/ and <https://github.com/KBNLresearch/Hackalod/tree/main/screenshots>
- ²⁶ Wikidata SPARQL query examples, https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples
- ²⁷ Creative Commons Attribution 4.0 International (CC-BY 4.0) license, <https://creativecommons.org/licenses/by/4.0/>