Abstract: How can one best transform a paper-based publication into a living online resource? This is the theme of a project at The National Museum of Art, Architecture and Design in Norway, supported by the Arts Council Norway. The National Museum aims to create, publish and maintain an authority list of Norwegian artists, architects, designers and craftsmen. The objective is to ease the digitisation process for other museums, scholars and the public in general and contribute to better data quality in Norwegian online collections. The list will in part be based on the Norsk Kunstnerleksikon (Encyclopaedia of Norwegian Artists in English), published in 1982–1986 and subsequently digitised in 2013. With the help of other public collections in Norway, the purpose is to make the new resource as complete as possible and available in both human- and machine-readable formats. Although the original paper publication contains biographical texts as well as lists of exhibitions, education, travels, publications and more, the data in the new authority list will be constrained to a set of core biographical data. It will however carry references to online biographical resources such as Norsk Kunstnerleksikon (NKL), Wikidata, Union List of Artist Names (ULAN) and Virtual International Authority File (VIAF). This article discusses the process of defining the scope of and setting constraints for the list, how to enrich and reconcile existing data, as well as strategies to ensure that other institutions contribute both as content publishers and end users. It will also shed light on issues concerning keeping such a resource updated and maintained.

Keywords: Name authorities; digitisation; from paper to digital; data cleaning; data reconciliation; interoperability; linked data

1. Introduction

From oral tradition to the great libraries of the ancient world, from Pliny’s Naturalis Historia to Wikipedia, people have sought to gather knowledge and craft constructs to disseminate it. The printed encyclopaedia is perhaps the quintessential manifestation of such an obsession. Yet, however educated and refined, our faculties have come up short when it comes to holding the wisdom of the world. A library, however vast and exhaustive, was inaccessible to most, but a book (or set of books) was something that could, at least in principle, be universally attainable. This was the realisation and fervour at the spread and industrialisation of the printing press. This ‘encyclopaedic vision’, as the historian Richard Yeo (2001) [1] has called it, was the idea that one could collate and create “a work containing the collective knowledge of a community” that could be spread, consulted and also, if the worst came to pass, serve as a “summary of intellectual accomplishments to be reactivated at a later stage . . . to be put together again if all other books were lost”). As shown in this quote by Denis Diderot, co-editor of one of the first modern encyclopaedias, this undertaking was viewed with the loftiest of ambitions:
“Indeed, the purpose of an encyclopaedia is to collect knowledge disseminated around the globe; to set forth its general system to the men with whom we live, and transmit it to those who will come after us, so that the work of preceding centuries will not become useless to the centuries to come; and so that our offspring, becoming better instructed, will at the same time become more virtuous and happy, and that we should not die without having rendered a service to the human race.”

(Diderot and d’Alembert, 2002) [2]

Diderot viewed his (and his co-editor d’Alembert’s) contribution as a “universal dictionary”, a distillation of all the world’s specialist and thematic dictionaries, lexica and vocabularies into one single work (albeit divided into seventeen volumes). The modern encyclopaedia would not only be a source of knowledge, but a hub of knowledge.

The production and spread of encyclopaedias, both universal and thematic, would continue with undiminished energy into the twentieth century, published in every conceivable language and relating to any imaginable topic. Most of these were, in the words of Richard Yeo (2001) [1], “seeking to record knowledge, not lives” (p. 16); they did not include history, people and biographies as this was the domain of a specific subtype of reference work—the historical, or biographical, dictionary. One such effort was the Norsk Biografisk Leksikon (Norwegian Biographical Encyclopaedia), which was published in nineteen volumes between 1923 and 1983 (with a revised second edition published between 1999 and 2005). In its preface it is stated that:

“The Norwegian Biographical Encyclopaedia shall contain biographies of all Norwegians that in one way or another have made themselves more than commonly known in the country—not merely in a single region—without consideration to whether or not their ventures have been deserving or not; there lies in the selection no other considerations than the biographical importance, none towards his ventures more or less fortunate or unfortunate character.”

(Bull, Krogvig and Gran 1923) [3]

This naturally also included the arts, but as an all-encompassing biographical reference work there was scarce space for lesser known or obscure artists, designers or architects. Also, as seen from the long timespan between the volumes in the first edition, many artists and craftsmen were excluded due to the timing of the period of their activity. The initiative thus arose to create a separate publication for Norwegian artists. Work commenced in 1977 as a joint effort by the National Gallery (Nasjonalgalleriet) and the Directorate for Cultural Heritage (Riksantikvaren) through funding from the Research Council (Norsk forskningsråd) and the Arts Council (Norsk Kulturråd). From 1982 to 1986, four volumes of the Norsk Kunstnerleksikon (Østby 1982–1986) [4], were published containing more than 3000 articles, long and short (See Figure 1).

The idea was that the encyclopaedia would be a “living” reference work that was continuously updated; thoughts of a digital edition were already present in the early 90’s. But as often happens, funding dried up and the involved institutions did not have the resources needed to keep the work going. Hence, the content was never updated, and revised editions never saw the light of day.

---

1 Original text: “... skal Norsk Biografisk Leksikon indeholde biografier av alle nordmænd som paa en eller anden maate har gjort sig mere almindelig kjendt i landet — altsaa ikke bare i et enkelt distrikt — uten hensyn til om virksomheten har været forfjentsfuld eller ej; der ligger i utvalget ingen anden vurdering end av den biografertes betydning, ikke av hans virksomhets mer eller mindre heldige eller uheldige karakter.”
That is, in a sense, until the digital revolution of the 2000’s. Norway was going digital, arts and culture had to follow, and funding was suddenly more available for projects that included digitisation. Funding was provided by the (now defunct) Archive, Library and Museum Authority\(^2\) (ABM-utvikling or Statens senter for arkiv, bibliotek og museum) and the Arts Council. The project was completed in two phases: First, from 2007–2012, work was done to transfer and transform all data from the paper source to a semantically structured machine-readable database, published as RDF (Data Nasjonalmuseet n.d.)\(^3\). When this work was concluded, it was decided that a human-readable web interface would be desirable as well and an agreement was reached with the online edition of the \textit{Store Norske Leksikon} (the Great Norwegian Encyclopaedia) (SNL, 2018) \(^7\) to publish the \textit{Norsk Kunstnerleksikon} within their own publishing framework. The online version (NKL.SNL n.d.) \(^8\) was unveiled in 2013. (See Figure 2 for example of interface). However, as others might have experienced, getting funding for a digitisation project with a defined scope is one thing, but getting funding for long-term, indefinite editorial work, maintenance and updates is another.

---

\(^2\) ABM’s mandate and resources were split between different governmental bodies.

\(^3\) A SPARQL endpoint is available (Data Nasjonalmuseet n.d.). This dataset is not updated.

\(^4\) Original text: “Å legge papirverk på nett er ikke en løsning på denne utfordringen, det er begynnelsen på en løsning. Nettpublisering i et rammeverk som muliggjør oppdatering av innhold er en forutsetning for arbeidet. Men innholdet krever like mye faglig og redaksjonell bearbeidelse som det trengte på papir for å være bra.”

---

\begin{figure}
\centering
\includegraphics[width=\textwidth]{norsk_kunstnerleksikon.jpg}
\caption{Norsk Kunstnerleksikon (Encyclopaedia of Norwegian Artists) 1982–1986.}
\end{figure}
instances, the information required is already readily accessible from other sources, but it must be collated. How then do you go about solving this problem, if a solution exists at all? How can you update and maintain an encyclopaedia, or any information really, when that requires the exact thing you do not have—resources?

In a way, this conundrum is nothing new; a century ago, with the creation of ever more editions and variants of encyclopaedias and the like, some came to conclude that in a way one was back to square one; with knowledge from differing fields, cultures and languages collected in different publications, how would one single out the most relevant ones? What was required was some sort of central, universally accessible deposit. The writer H.G. Wells (1938, 63) [10] stated it in this way:

Figure 2. Screenshot of the web interface of Norsk Kunstnerleksikon.
“It is dawning upon us, we lay observers, that this work of documentation and bibliography, is in fact nothing less than the beginning of a world brain, a common world brain. What you are making me realise is a sort of cerebrum for humanity, a cerebral cortex which (when it is fully developed) will constitute a memory and a perception of current reality for the entire human race. Plainly we have to make it a centralised and uniform organisation but . . . it need not have any single local habitation . . . In these days of destruction, violence and general insecurity, it is comforting to think that the brain of mankind . . . can exist in numerous replicas throughout the world.”

In this day and age of the internet, we have attempted to create this “world brain”: Perhaps not centralised and uniform, but more infinite than even Wells could have conceived of. Herein lies the challenge; not of producing content—the world brain already has the content—but of structuring that content and being able to retrieve it at the right time. Accordingly, we will not be creating a new edition of the Norsk Kunstnerleksikon; rather, we will create an authority list of artists, architects, designers and craftsmen present in Norwegian public arts collections, an index of names linked to further information elsewhere. With the help of other public collections in Norway, we aim to make the new resource as complete as possible and available in both human- and machine-readable formats. Although the original paper publication contains biographical texts as well as lists of exhibitions, education, travels, publications and more, the data in the new authority list will be constrained to a set of core biographical data. It will instead carry references to online biographical resources. It will be a connection of open information from disparate sources linked through name authorities. For this task, funding has again been provided by the Arts Council Norway.

2. Background

Over the last decade, the National Museum has invested considerable effort into improving the digital management of its art collection information. Working in the professional area of digital collection management, the National Museum aims to provide consistency, clarity and context for the digital catalogue, and is driven by the intention to ‘Create Once Publish Everywhere’ (COPE)[11]. Like many other museums in the 2010s, information specialists at the National Museum worked on specifying new conditions for digital information and analysing new types of metadata and layers of structure. During different digitising projects, the National Museum benefited from best practices and deliverables from others, such as the CIDOC committee’s standards and guidelines, the Getty’s Online Scholarly Catalogue Initiative or EU Program-funded projects such as Digitising Contemporary Art (DCA) and Terminologies as a Service (TaaS) [12].

In 2012, the Arts Council Norway funded the national terminology platform for cultural heritage, KulturNav (kulturnav.org), initiated by the software company KulturIT®. KulturNav aimed to facilitate the management and delivery of shared terminology and thereby paved the way for establishing new online authority lists in the Norwegian museum sector. Together with KulturIT and the art museum community, The National Museum started working on standardising art history terminology such as object types, materials and techniques, linking as much as possible to online resources (such as The Art and Architecture Thesaurus) and publishing the lists on the web as authorities. Regarding artist names, one of the fundamental provenance information elements is the creator identity[7][13].

---

5 A concept strategy made well known to the museum community by Nick Poole, as chair of the Collections Trust in Britain. Poole (2016) describes the maxim as born—“coined in a blog post by Daniel Jacobson, Director of Application Development for National Public Radio in the US—and some powerful new work being done by the British Museum as part of their ResearchSpace project.”

6 KulturIT is a provider of a collection management system and other IT infrastructure for museums in Norway and Sweden. Originally a unit within the public ALM authorities in Norway, it was later transformed into a private company and is now owned by a consortium of public museums in Norway and Sweden. Product development and services are funded by the shareholders, licensing fees and to a significant extent by funding from Arts Council Norway.

7 Ref. Categories for the Description of Works of Art (CDWA) ‘4.1.3 Creator Identity’ [13].
The National Museum had no more than linked the names of foreign (non-Norwegian) painters and sculptors in the collection to a Virtual International Authority File (VIAF) [14] and created an authority list of the museum’s preferred spelling of them. Throughout the museum’s digitisation projects (2008–2018), the task of establishing a creator identity authority list for Norwegian artists was put on hold. The museum waited for resources and better technical solutions and hoped for the important work of the scholarlarly prestigious Norsk Kunstnerleksikon to be continued. Until the start of this project there was no up-to-date national or internal authority list of identities to use for the digitised works of artists, architects, designers and craftsmen in the collections. The artist encyclopaedia has been the only authority list of Norwegian artists names, even though the museum never technically integrated it with the collections management system.

Starting the project, the National Museum wanted to concentrate on areas not covered by the Norsk Kunstnerleksikon. Since the encyclopaedia has never been linked to other name identity authorities, we wanted to link all identities, both coming from the encyclopaedia and from this project, to already established authorities such as the Union List of Artist Names (ULAN) [15] and VIAF. During this decade of emphasising the digitisation and publishing of the collections, the expectations regarding scope, function and format of a list of artist names has changed in various ways.

3. Scope

How to define the scope of an authority list? As a starting point, the National Museum had a list of more than 28,000 documented name entries in the collection management system. This was a list of variable quality, inconsistent in structure and with many duplicates. The names represented persons, companies, institutions, factories and (even) mythological figures, and was the accumulated result of twenty years of cataloguing. We needed to streamline the content and decide which information elements to include and how to define them. The concept Artist is an example of a term defined in variable ways. Our definition: ‘any person with a known name who has created works collected by The National Museum of Art, Architecture and Design in Norway’ is slightly different from, e.g., ULAN’s definition. In comparison to ULAN, we excluded workshops, families, groups of artists, but also anonymous and unknown artists. Entries where we could not identify a proper name were also left out. Further on, we also excluded companies, groups and institutions (hereafter corporate entities) to focus on a uniform data model that could be applied to all entries. We had a year to complete the project with resources of one full-time equivalent employee, and to work with a list of corporate entity names would be too time-consuming.

Our person list comprised approximately 14,000 records. It soon became apparent that concentrating on Norwegian artists was counterproductive. It made sense to exclude the criteria ‘Norwegian’ for the sake of interoperability with other published authoritative name lists in the future, and we expanded the scope to include all artists in the collection. This way, one can use the preferred spelling of all names, and within a longer timeframe contribute to making other authority lists more complete. We believe that maintaining a list of artists represented in the collections will be useful for other third-party users both nationally and internationally and for the sake of the museum’s own cataloguing purposes it makes perfect sense. For those specifically requiring it, an authority list of only Norwegian artists would still be available as a subset of the full list. Or a list of French painters in our collection, for that matter.

8 We use Corporate entity in the same meaning as Corporate body in Cambridge dictionary: ‘an organisation such as a company or government that is considered to have its own legal rights and responsibilities. https://dictionary.cambridge.org/dictionary/english/body-corporate. We thereby exclude persons from the definition.
Deciding on a reduced set of core elements was challenging, and due to limited resources and a short time frame we ended up with less information than initially intended. We had to prioritise the production of an awaited resource with core information elements that could provide the basis for linking to both additional biographical information and any other desirable or necessary information in the future. At the same time, even core elements related to a person’s identity acquires decision-making about relationships, associative relationships, preferred terms, pseudonyms and uncertain information. Our data quality analysis then resulted in the exclusion of relationships and some of the uncertain information. This resulted in a minimal set of fields (see field list in Figure 3) compared to the core elements recommended by the managing editor at the Getty Vocabulary Program, Patricia Harpring (2010, 102) [16].

This minimal field set identifies the artists and makes it possible to differentiate them from each other. Gender was added as a core element since this plays a role in the museum’s annual governmental reports (e.g., how many works by female or male artists are acquired or exhibited per year, and so on). Place of birth and death have also been maintained as fields in the collection management system since these are conventional parts of exhibition labels in the National Museum.

To structure the information elements, we then added editorial rules (e.g., required, optional and repeatable fields) and established rules for how to choose a preferred term and which variant term to include (e.g., to use the transliteration as found in the most authoritative of available sources). We standardised format, syntax and character sets and discussed available authorised sources for each field. We decided to depict ambiguous information in repeatable fields, with one preferred entry and other possible values for name spellings. Uncertain bibliographic information, e.g., birth date, was excluded since it had not been documented in the conventional manner (circa/possibly/either-or). Support for this was unfortunately lacking in the original data structure, and we had no time to add this.

**Figure 3.** Core elements of Norwegian Artist Names Authority List of Artists in Norwegian Art Collections.

### 3.1. Method and Tools

Knowing what elements to include, we could then export them. The data from our collection management system could be retrieved with an SQL query that joined the name table with data from other tables, such as roles, and had checks such as ensuring the persons were creators of objects in the collection. The results of this query were then exported to a flat list in Microsoft Excel. We kept the data unchanged in our database, both as a security save but also to be able to keep working in the database during the project. In our spreadsheet of artists, each row consisted of one person and each column had one type of information about the person. Some of these columns were easily defined, others not.
One example is the life role column. The collection comprises fine art, architectural, design and craft objects collected throughout our institution’s history. Hence, not all persons in our list are artists by profession. In similar publications on Kulturnav.org, the persons are all identified with “artist” as a life role. We wanted to define this further and specify whether they were painters, architects, designers, engraves, etc.: Firstly, because our dataset already had this kind of detail on life roles, it would not make sense to lose this information, and secondly it could prove useful in the process of identifying the right person in other web resources such as VIAF, ULAN and Wikidata [17]. Records with similar names could be checked for a match against other sources by including information such as birth or death date and the profession column to clarify the correspondence of the two records. We allowed for multiple values in the life role field to give a fuller biographical overview. As an example of multiple life roles, we have authors who are the author of a text but also the painter of the illustrations in a book; this person would then be called author and painter in our list.

After determining which columns to include about each artist, we started editing the list in the data transformation application OpenRefine9 [18]. OpenRefine is free, open-source software that can be downloaded and run locally. This program was a perfect fit for us. By using the very powerful sorting, filtering and facet tools, we could get an overview of our dataset so far and start identifying problem areas that would need special attention. An example was the column nationality: this information was from our collection management system a free text field. This meant a lot of inconsistent spelling. By using text facets, we could easily correct the errors (see Figure 4). This method was used in almost every column to check and correct spelling.

![Figure 4](image-url)

**Figure 4.** A screenshot showing how the text facets were ordered by count to see how many artists were from each country, and also showing where we cleared up errors such as wrong spelling and missing capital first letter.

By using the text facets, we could also cluster similar cells—e.g., if the column included cells that had the values ‘Canada’, ‘canada’ and ‘cannada’, OpenRefine’s cluster tool identified these as similar and we could change them all to what we preferred (‘Canada’) (see Figure 5). After using these tools in every column, the list gradually became more complete and consistent. We could now search for cells with missing values.

---

9 OpenRefine—formerly Google Refine—is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data. (OpenRefine n.d.) [19]
3.1.1. Reconciling our Data with Other Online Resources

To start filling in missing values, we reconciled our data with other online resources, namely viaf.org, norskkunstnerleksikon.no, kulturnav.org and wikidata.org. For both VIAF and Wikidata there are existing reconciliation web services that can be queried from within OpenRefine. We decided that matching against external web resources would only be carried out automatically, by automated query processes using the reconciliation tool in OpenRefine, and that any manual mapping was outside the scope of our project. Starting with viaf.org, the search found a lot of matches to our person records, but some were uncertain matches. We decided on how high the score had to be to automatically link our record to a VIAF record; the ones with scores too low were manually checked for correctness. For about 50% of our person records, we were able to create links to web resources. Some persons were matched to several resources and others just to one. This taught us that it is important to link to different kinds of web resources. We saw that some of the web resources had many more Norwegian artists than others and that some of the resources had more detailed information about international artists. The spelling of names also made the linking a bit more troublesome as some resources would only match if there was exact correspondence on spelling. The Norwegian letters ø, æ, and å also produced some difficulties with matching the artist names to external resources, but for most resources this was solved with troubleshooting string encoding. For the other resources we had to spend some time working out the queries to send to the application programming interfaces (APIs) of the different web services. The SPARQL endpoint for the RDF dataset of the Norsk Kunstnerleksikon was particularly tricky, passing a very lengthy URL-encoded SPARQL query to the endpoint through the GREL syntax of OpenRefine’s reconciliation tool.

---

10 Some of OpenRefine’s functionality requires defining queries, value comparisons and more using GREL (Github 2015) [19].
3.1.2. Enriching Our Dataset with Information from the Linked Records

Having linked as many records as possible, we could now harvest data from the links. We used wikidata.org to collect core information such as birth, death, gender and nationality. This would complete the list even further, but also worked as a cross-check to see if our information corresponded with the info from Wikidata. External data was initially stored in separate columns to enable automated column comparisons with GREL. This was particularly useful for birth and death dates. In a great number of cases we were able to improve a year value in our original dataset with a full date (D-M-Y) from Wikidata. However, some data did not correspond and we had to do an extra check to see if it was the same person and then either correct our own data or ignore the data from Wikidata. This was done based on a search on which data we could find in other sources and whether the same data occurred across multiple sources. This meant that if the data from Wikidata was only found in Wikidata and our “original” metadata was found on several websites, we would keep our “original” info and discard the info from Wikidata but keep the link. In some cases, there would be discrepancies, for instance in birth and death year, that were small enough to indicate that these records still referred to the same persons. Such minor discrepancies occurred almost exclusively for years before the twentieth century and, not surprisingly, more frequently the further back in time. The differences could be identified automatically by setting a tolerance of, e.g., +/- 2 years for date matching, then manually evaluating and addressing the suggested matches.

After we had gathered core elements from Wikidata, we found an error in some of the dates. Suspiciously many people were born or died on 22 December during the sixteenth century. After some additional searching to find these dates, we had to conclude that some data corruption had occurred in the process of enriching years with day and month from Wikidata. We decided to remove these dates from our list. In this vast dataset we had to make some pragmatic choices to complete it, even if this resulted in removing some good data. All the artists born in the sixteenth century now have no calendar date of birth or death, just the year that was already in our original dataset.

As we kept supplementing our dataset with a growing number of links to matching entries in external web resources, it became clear to us that finding these links was just as important as finding exact and correct bits of information to enrich our own core data. This was at first somewhat troubling to us as museum workers, but we realised that the links have great future potential to provide more information on a person. The link was just as important as making sure that every bit of information we collected was correct.

3.1.3. When Do We Stop?

Filling in the blank cells was a time-consuming process, but a very important one. Many of the artists are not found in the different online resources but can be found on other websites. This is the case for many Norwegian contemporary artists, but also many of the older Norwegian artists. Not surprisingly, we were unable to find information on all persons. At some point we had to stop and accept that there will be many records with some missing data elements. The authority list, when finished, consists of almost 10,000 people records. (See Figure 6 for illustrations). As we aim to make this a collaboration between museums in Norway, we hope that others will contribute later and fill in the blanks where they happen to have information. Regional museums may have more detailed information on some of their local artists than what we have found. The long-term success of the project will depend on collaboration between museums and willingness to share information with each other.
At the same time as we worked on narrowing our scope of content to a minimum, the possibilities concerning options for technical structure and web interoperability seemed to increase. To be able to decide on this, we first had to make a clear definition of the intended function(s) of this list.

We have had the purpose of our authority list in the back of our minds throughout the years and have had to revise our view several times. Our list of names was initially made by, and only meant for, cataloguers internally at the museum. After publishing the collections on the national cultural heritage portal Digitaltmuseum.no in 2010, the list had to be made available at least to all cataloguers in Norwegian museums who wanted to share content on this portal. In the following years, the user demand for an authority name list as a web resource increased proportionately with the growing number of collections being published online. As we then took up this project again in 2017, it was a long-established fact that authority lists like these are created for retrievers across different types of collections and continents and for referencing semantic web technology.

4.1. Fit for Purpose?

The National Museum values scholarly conventional, updated and reliable information. As the responsible museum for this list, we want, as in the “lessons learned” of the OSCI, (the Getty Foundation 2017, Lesson 1) [20], to underline the importance of our content standards to be trusted also in the digital environment. To reiterate our rationalised scope: We now have quality controlled

Figure 6. Graphical breakdown of the mentioned results.
4. Function

At the same time as we worked on narrowing our scope of content to a minimum, the possibilities concerning options for technical structure and web interoperability seemed to increase. To be able to decide on this, we first had to make a clear definition of the intended function(s) of this list. We have had the purpose of our authority list in the back of our minds throughout the years and have had to revise our view several times. Our list of names was initially made by, and only meant for, cataloguers internally at the museum. After publishing the collections on the national cultural heritage portal Digitaltmuseum.no in 2010, the list had to be made available at least to all cataloguers in Norwegian museums who wanted to share content on this portal. In the following years, the user demand for an authority name list as a web resource increased proportionately with the growing number of collections being published online. As we then took up this project again in 2017, it was a long-established fact that authority lists like these are created for retrievers across different types of collections and continents and for referencing semantic web technology.

4.1. Fit for Purpose?

The National Museum values scholarly conventional, updated and reliable information. As the responsible museum for this list, we want, as in the “lessons learned” of the OSCI, (the Getty Foundation 2017, Lesson 1) [20], to underline the importance of our content standards to be trusted also in the digital environment. To reiterate our rationalised scope: We now have quality controlled descriptive metadata in a list shaped by the character of our collections’ content. It is categorised by international standards but also by organisational preferences (such as the information units: Gender, birth/death place).

For the authority list to contribute to improved knowledge about Norwegian art, or art in Norwegian collections, one must consider a variety of formative steps. The question of whether the data is good enough, for further use, appeared once again; with the COPE-maxim in mind: Are the data fit for reuse in 2018? We experienced what information scientists Seth Van Hooland and Ruben Verborgh state in Linked Data for Libraries, Archives and Museums: “it is sometimes extremely complex to find consensus on whether data are of good or bad quality” (2014, 73) [22,23]. Several scholars have over the last years analysed the quality of data and the utility of web services within the area of cultural heritage. Concerning an authority list as a web resource, we had two main issues that we focused on in our further examination of quality and usefulness.

The first issue concerned the narrow scope of our descriptive metadata. As associate professor of visual culture Nina Lager Vestberg (2013) [24] points out: “… an online database may just as easily close off avenues of inquiry as open them up, thereby limiting rather than expanding the number of potential questions to be asked of the archive” (487). She analyses how archival order ‘affect[s] the production of knowledge and meaning’, by comparing methods used in digitising web archives in the photographic collections of the Warburg Institute and the Courtauld Institute of Art in London. Vestberg exemplifies how sticking to the conventional cataloguing structural manner might exclude the expanded audience that the web service could have been intended for: “it should come as no surprise that an image resource designed by iconographers will tend to be of use chiefly to other iconographers” (484). Hence, in our case, adding descriptive metadata to the authority list was already out of our scope. We wanted to make the list available for linking to further content, and therefore went on evaluating how to structure it as linked data. The second issue thereby targeted the more technical side of interoperability, as we concentrated on adding layers of metadata encoding to the authority list. Strategies for implementing linked data are published by CIDOC’s Statement on Linked Data Identifiers

---

11 Descriptive metadata, meaning content in the definition from DCA: “Descriptive metadata is used to identify and describe collections and related information resources, e.g., artist, title, location, date [. . . ]” (DCA 2012, 8) [21] and based on the CDWA standard (Baca and Harpring 2009) [13].
for Museum Objects (2013) [25]. Suggesting answers and practical guiding to the challenge of reuse, also in the sense of machine-readable formats, is done, e.g., by the abovementioned Hooland and Verborgh (2014). They provide expedient explanations on how quality metadata can be applied to existing descriptive metadata and how this is key to better access to content.

4.2. Question of Format

Having these issues in mind, and being aware of how rapidly data formats are evolving around us: How can one make good choices to meet the expectations of a name authority to be available both for humans and machines to link, systematise and enrich? We concluded that the importance of achieving interoperability with international standards is of greater importance than ever. To achieve interoperability, we found that it was decisive to apply editorial and structural rules of standard vocabularies where possible. Unlike the ULAN, our authority list is not a thesaurus. It only has one level and all terms belong to the same class (person). Our list nevertheless includes same-as-references (links to VIAF, ULAN, etc.) and preferred terms (name spelling, nationality and life roles). Considering the topics of technical format/ metadata encoding, we aspire to add: multi-language support for all text fields, controlled text, date and numeric formats and stable URI/ permalinks. We also hope to map it to a concept reference model, preferably CIDOC CRM [26]. Once published to KulturNav, the artist list will be available for use in the KulturNav API as RDF. KulturNav’s framework has a number of entities, one of which is the person entity with the following data structure (see Figure 7), which covers our chosen data elements:

![Figure 7. Used KulturNav data elements.](image)

KulturNav’s provider KulturIT is currently working on improved and extended use of standards such as CIDOC-CRM, DCAT-AP and Schema.org (Kulturnav 2018 n.d.) [27]. Metadata expert Tony Gill (2016) explains in Metadata and the Web, the different data structure, -format and -value standards, tools and methods for encoding metadata in a machine-readable form, and protocols for distributed search and metadata harvesting. He argues that: “By using these various components in intelligent and appropriate ways in order to provide access to the rich information content generated by libraries, archives, and museums, it should become possible to build a global Semantic Web of digital cultural content and integrated search tools to help users find the content they are seeking” (Gill, 2016) [28].

---

12 “The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the “semantic glue” needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.” (CIDOC CRM n.d.) [26].
In this way, Wells’ concept of the world brain could potentially be realised. These museum community recommendations also harmonise with the new guidelines for public data availability published by the Norwegian Agency for Public Management and eGovernment (DIFI), where the following is one of their highlighted points:

“Data must be available in machine readable formats. In addition, the formats should be standardised. This ensures good interactivity with other information (interoperability) and does not impose unnecessary limitations on what the information can be used for in the future. Examples of machine readable and standardised formats are CSV, XML, JSON and RDF serialisations such as RDF/XML, JSON-LD, and Turtle."\(^\text{13}\) (DIFI 2018, subsection 8) \(^{[29]}\)

The National Museums goal in standardising all authority list information components is to achieve sustainability and be equipped to adjust to new evolving data formats in the future. However, some decisions about sustainability, richness and progress remain to be made for the future of the authority list.

5. Future Plans

Management and Further Plans

As we have seen, the major shortcoming of *Norsk Kunstnerleksikon* is that it is not a living resource. At the National Museum in Oslo, we have been using the online RDF version of the encyclopaedia to harvest and display biographies of artists in our collection online \(^{[30]}\). We are seeing more and more that a number of biographies do not represent the individual artists well enough, and in many cases more recently updated biographies can be found elsewhere, such as on Wikipedia or in the online version of the *Store Norske Leksikon*. An online artist list resource with links to external resources is a great tool for harvesting the best biographies—or even giving the public the choice between different sources, like the Museum of Modern Art in New York does \(^{[31]}\).

Our authority list of artists needs to be managed and edited in the future so that it is a living resource. One way of keeping the information updated could be to automatically harvest new or revised data through the links to external resources. This could be time saving for the museum, but we have still not decided on whether to do this. One of the problems with automatically changing the information is of course adding errors and that correct information gets “lost”. This is a discussion that we have not completed yet. A more likely scenario is an extension of the joint museum community effort. The National Museum intends to involve and encourage other national art, design and architecture collections to make use of and contribute to keeping the list updated and supplementing it, through the editorial interface and workflow of KulturNav.org. In addition to this, one could envisage a long-term and semi-automated crowdsourcing project where the public are encouraged to contribute through an interface in the National Museum’s online collection (or other museums’ online collections or Digitaltmuseum.no). Through an interface, the public would be able to report errors, supply missing information, report unrecorded deaths, and so on. These submissions can find their way to the editorial workflow of KulturNav for manual checking and submission by the editorial team.

As a start for future sustainability, at the National Museum in Norway we commence to take on the editorial task and will encourage the Norwegian museum community network to participate in

---


\(^{14}\) Nasjonalmuseets collection contains around 400,000 art, architecture and design objects. More than 42,000 objects are currently available online, and the number increases as we publish new objects every month.
keeping the authority list up to date. We have established an editorial team, and hope that before the authority list is published people representing the triangle of finances (the Arts Council and other financing sources), infrastructure (KulturNav and IT providers) and content (collection cataloguers, the public) have committed to a strategy. If we manage, we would ensure that the authority list can live on and evolve to cover the long-awaited authority list that can ensure consistency, clarity and context in a professional way for digitised Norwegian art.

KulturNav, as a platform for shared authority lists and terminology in Norwegian and Swedish cultural heritage, has been slowly gaining momentum over the last couple of years. It has taken time for end users in the museums to see how such a platform can interoperate with other museum tools such as collection management systems, online collections and similar, and how it can provide value to both internal workflows and data quality. It has also taken time to establish an understanding among interested users that the success of any such platform must be a community effort. There is a huge potential for shared authority lists in Norwegian museums. So far, many lists have been smaller local initiatives with a clearly defined and limited scope, for instance a list of designers working for a particular porcelain factory. As our authority list of artists in the National Museum of Art is published to KulturNav, we plan to organise several local workshops for the Norwegian museum sector to engage other museums in the effort to improve and contribute to the list. Such workshops will not only evangelise about the benefits of a joint authority list for artists, but must also provide detailed training on how to contribute, editorial workflows and policy.

Author Contributions: Conceptualization, Methodology Writing—Original Draft Preparation and Writing—Review & Editing by Y.B.H., D.H., G.B.P. and M.B., Formal Analysis and Investigation by Y.B.H.; Project Administration and Funding Acquisition by D.H.

Funding: This research was funded by Arts council Norway, grant confirmed 13.01.2017.

Conflicts of Interest: The authors declare no conflict of interest.

References


