

# Fascinating with Open Data: openArtBrowser\*

Bernhard G. Humm<sup>[0000-0001-7805-1981]</sup>

Hochschule Darmstadt – University of Applied Sciences  
Haardtring 100, 64295 Darmstadt, Germany  
bernhard.humm@h-da.de

**Abstract.** This in-use-paper presents openArtBrowser, a Web application for educating in visual art, fascinating users for paintings, drawings and sculptures. OpenArtBrowser is solely based on linked open data and its code is open source. It fosters serendipity by supporting users to discover new aspects of art out of curiosity, without actively searching. The user interaction concept and software architecture is explained and discussed.

**Keywords:** Linked open data, semantic search, serendipity.

## 1 Introduction

In this in-use-paper we present *openArtBrowser*<sup>1</sup>, a Web application for educating in visual art, fascinating users for paintings, drawings and sculptures. OpenArtBrowser is solely based on linked open data, namely from *Wikidata*<sup>2</sup>, and Wikimedia Commons<sup>3</sup>, and its source code is open source<sup>4</sup>.

This work has partly been inspired by our earlier works in *digital heritage* and *GLAM (galleries, libraries, archives and museums)* [1-6] which resulted, amongst others, in the digital collection of Städel Museum Frankfurt<sup>5</sup>, one of the most prominent art museums in Germany.

OpenArtBrowser serves two *use cases*:

1. *Active search*: The user wants to retrieve specific information about visual art, e.g., details about a painting.
2. *Browsing*: The user has no need for specific information but wants to be inspired by art, e.g., looking at artworks that fascinate him or her and learn interesting aspects.

---

\* Copyright © 2020 for this paper by its author. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<sup>1</sup> [openartbrowser.org](http://openartbrowser.org)

<sup>2</sup> [www.wikidata.org](http://www.wikidata.org)

<sup>3</sup> [commons.wikimedia.org](https://commons.wikimedia.org)

<sup>4</sup> [github.com/hochschule-darmstadt/openartbrowser](https://github.com/hochschule-darmstadt/openartbrowser)

<sup>5</sup> [sammlung.staedelmuseum.de](http://sammlung.staedelmuseum.de)

With openArtBrowser, we pursue the following *goals*:

1. *Learning with fun*: The user shall learn new aspects of visual art in a joyful and playful manner.
2. *Open Data*: All data about visual art shall be from open sources.
3. *Serendipity*: The user shall discover new aspects of art out of curiosity, without actively searching for it. Also regular users shall be surprised by unexpected findings over and over again.
4. *Usability*: The application shall be easy to use, i.e., be simple, consistent and self-explanatory, shall focus on relevant information, have low response time etc.
5. *Aesthetics*: Corresponding to the domain of visual art, the application shall convey an aesthetic appearance.
6. *Responsive Design*: The application shall be usable on various devices, including desktop computers, tablet computers, and smartphones.

## 2 User Interaction Concept

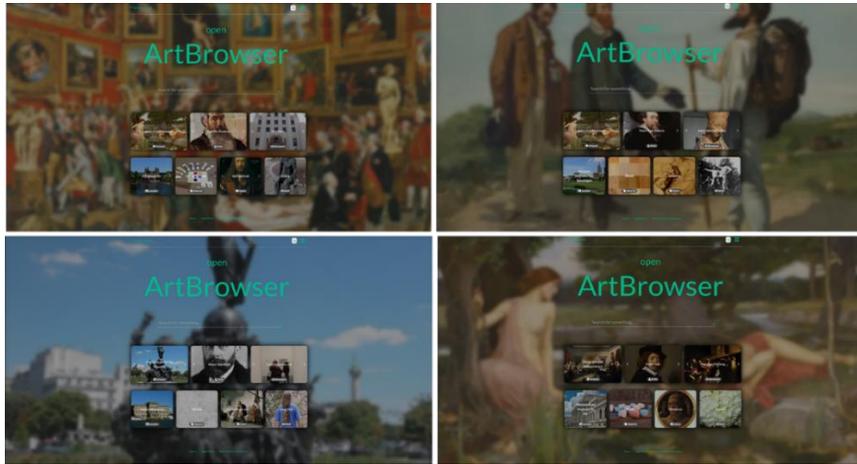
In this section, we introduce the user interaction concept of openArtBrowser by presenting some of the pages.

### 2.1 Homepage

The homepage of openArtBrowser provides two prominent elements:

1. A *search bar* allows actively searching for visual arts. Animated suggestions like “Try Mona Lisa” invite to start searching.
2. Seven tiles with *search facets* show suggestions for concrete artworks, artists, movements, locations, materials, genres, and motifs. The suggestions are animated and change from time to time.

The home page changes appearance each time the page is visited or re-loaded. See **Fig. 1**.



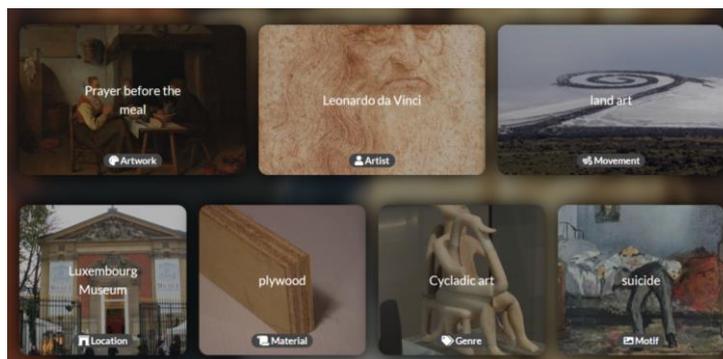
**Fig. 1.** Different appearances of the openArtBrowser homepage

## 2.2 Facets

The following search facets are provided:

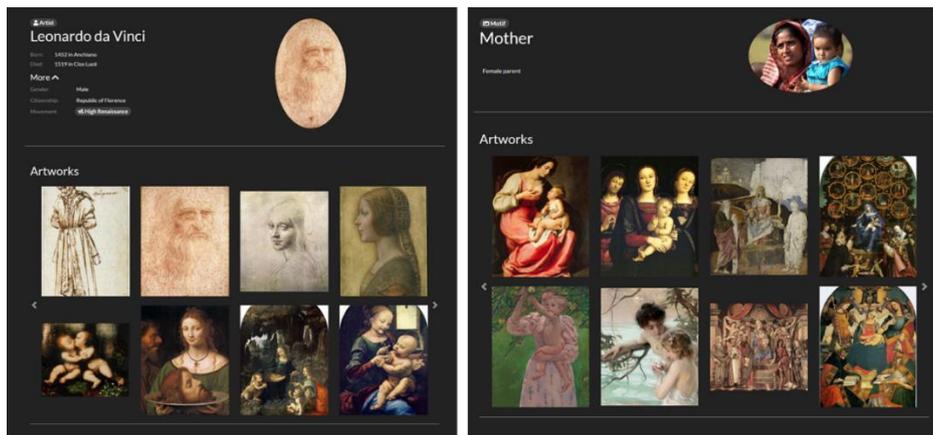
1. *Artwork*: Concrete paintings, drawings, or sculptures, e.g., “Prayer before the meal” by Quirijn van Brekelenkam
2. *Artist*: painter or sculptor, e.g., Leonardo da Vinci
3. *Movement*: Artistic movement, e.g., land art
4. *Location*: Places where artworks are exhibited, e.g., Luxembourg Museum
5. *Material*: Materials that were used in artworks, e.g., plywood
6. *Genre*: Artistic genres, e.g., Cycladic art
7. *Motif*: Things or aspects depicted in an artwork, e.g., suicide.

Each facet has a unique icon which is used consistently throughout the application.  
See **Fig. 2**.



**Fig. 2.** Search facets on the homepage

For each facet there are pages with interesting details. For example, on artist pages the birth dates and death dates and of artists, their citizenship, artistic movements etc. are displayed. In the lower part of the pages, artworks are shown which match the facet. E.g., on the page for the motif “mother”, depictions of mothers are shown. Sliders allow to show more artworks which do not fit on one page. An animation moves the sliders from time to time. See **Fig. 3**.



**Fig. 3.** Examples of artist page (Leonardo da Vinci) and motif page (mother)

### 2.3 Artworks

The focus of openArtBrowser is on artworks. When selecting an artwork in a facet page, then the respective artwork page is opened. See **Fig. 4** for an example. Depicted is the upper part of the artwork page. The lower part (related artworks, **Fig. 5**) is described below.

The upper part of the artwork page contains a depiction of the artwork itself and metadata. When expanding the image in full-screen mode, a high resolution image is loaded which can be zoomed to inspect details.

The metadata contains details about the artist, the location, and the inception of the artwork. To focus on the painting itself, more metadata are hidden when opening an artwork page but can be displayed with one click as shown in **Fig. 4**.

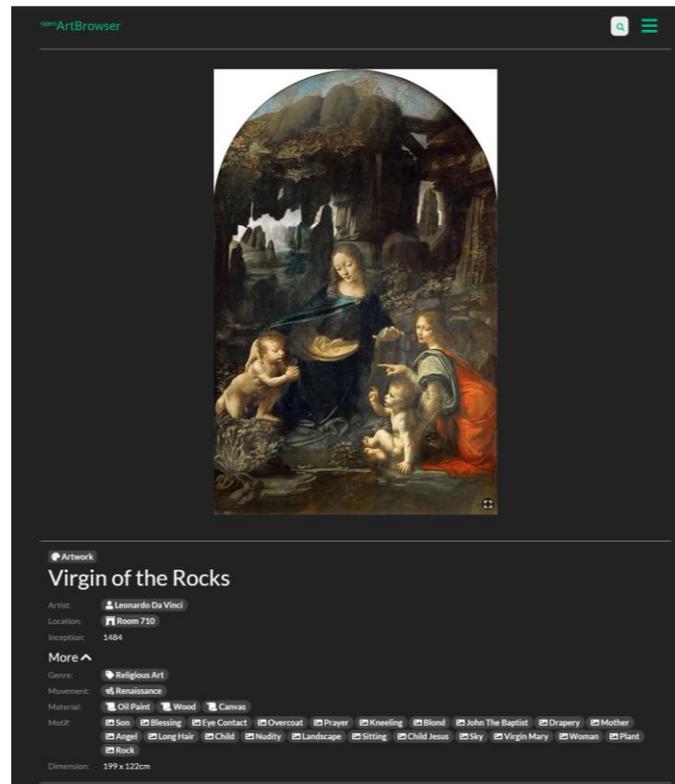


Fig. 4. Example of an artwork page (Virgin of the Rocks by Leonardo da Vinci), upper part

## 2.4 Related Artworks

At the bottom of an artwork page, related artworks are depicted. Artworks are considered related if they share at least on motif, artist, location, genre, movement or material. See Fig. 5.

In different tabs, related artworks for one specific facet can be displayed. The tab “All” which is activated when opening an artwork page combines all facets. When hovering over one of the related artworks (in Fig. 5 the artwork “Madonna and Sleeping Child with Three Angels”), then those tags are highlighted which are shared between both artworks. E.g., The artworks “Virgin of the Rocks” and “Madonna and Sleeping Child with Three Angels” both are oil paintings on canvas, belong to the genre religious art and depict virgin Mary, child Jesus and angels, and are both exhibited in Room 710 of the Louvre Museum in Paris. So, using this feature, different artworks can be compared.

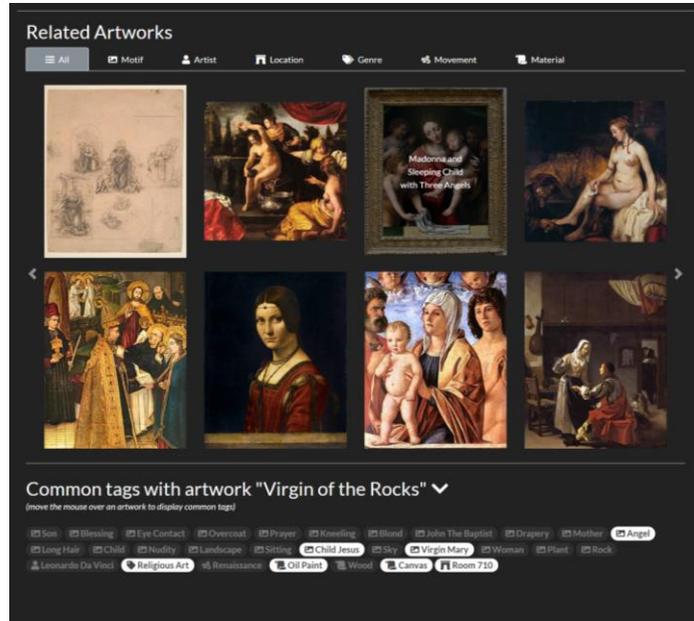


Fig. 5. Related artworks (artwork page, lower part)

## 2.5 Semantic Search

In all pages of the Web application, the search bar can be used. While the user is typing search terms, search suggestions are displayed, similarly to a Google search. Unlike the Google suggestions, the openArtBrowser suggestions are disambiguated semantically and assigned to the search facets. See Fig. 6.

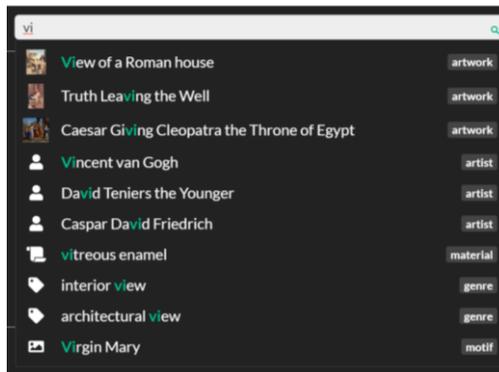


Fig. 6. Search bar with semantic autosuggest when typing “vi...”

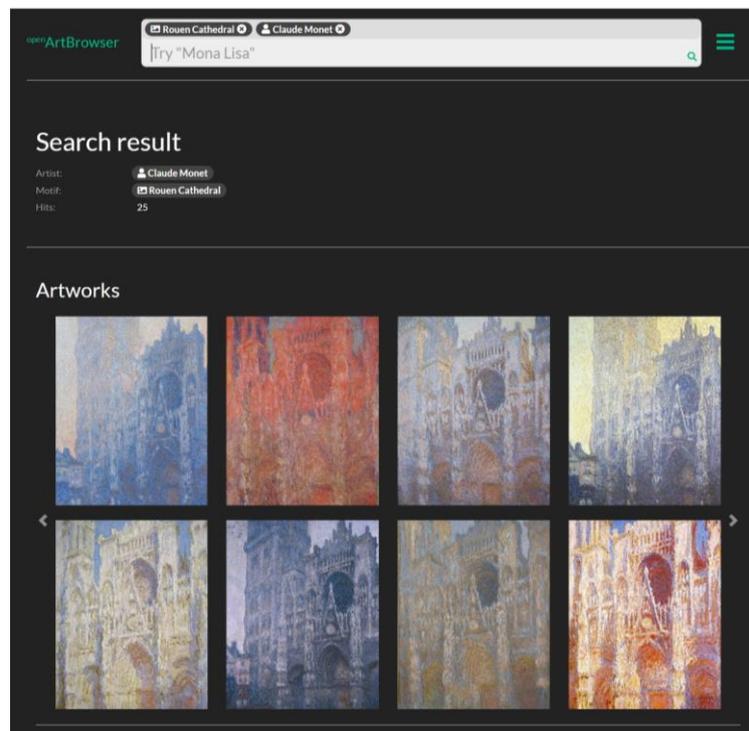
In the example of Fig. 6, the user is typing the letters “vi...”. Various artworks, artists, materials, genres and motifs are displayed which contain the letters “vi” (case-

insensitive), grouped according to their facet. The matching letters “vi” are highlighted (in green). A sophisticated heuristic ranking selects a limited number (here 10) of suggestions from a potentially very large number of matches. Ranking criteria include:

1. *Ranking of artworks and facets*: The ranking heuristics include: The more information available for an artwork the higher its rank; the more artworks exist for a facet the higher its rank.
2. *Position of match*: The ranking heuristics include: Matches at the beginning of the first word (e.g., “Virgin Mary”) are ranked higher than matches at the beginning of another word (e.g., “architectural view”), which are, again, ranked higher than matches within words (e.g., “David”).
3. *Diversity*: More different facets are favored over only one or very few facets, even if their ranking according to criteria (1) and (2) might be lower.

The concepts of the semantic autosuggest feature are oriented at [7].

Whenever a user selects an autosuggested term, the respected page is opened immediately. However, the user may continue to refine the search by adding more search criteria. See **Fig. 7**.



**Fig. 7.** Multiple search conditions

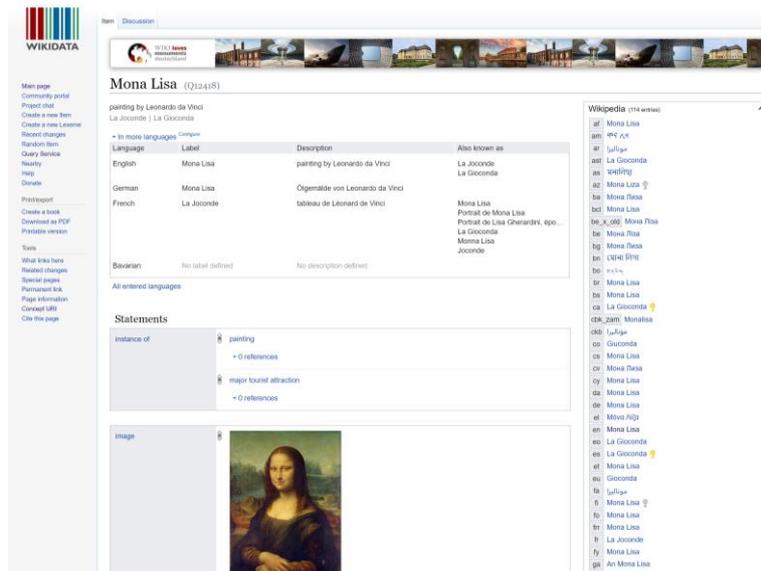
In the example of **Fig. 7**, the user has specified the search criteria “Rouen Cathedral” (motif) and “Claude Monet” (artist). If more than one search condition is entered, a search result page is opened that shows artworks which satisfy all search conditions (AND connected search); in the example all artworks by Claude Monet that depict Rouen Cathedral (25 hits on display).

When the user enters a search string but does not select any suggestion, a full-text search in the labels of all entities is performed. A full-text search may also be refined by additional search criteria.

### 3 Linked Open Data Source

All data displayed in openArtBrowser is open source. All metadata stems from *Wikidata*<sup>6</sup>, a collaboratively edited knowledge base hosted by the Wikimedia Foundation which is used, e.g., by Wikipedia. All images are hosted at *Wikimedia Commons*<sup>7</sup> under a Creative Commons License.

Wikidata is a *knowledge base* (a.k.a. *knowledge graph, ontology*) that can be read and edited by both humans and machines. It stores topics, concepts or objects, their attributes and relationships. See **Fig. 8** showing the first page of the Wikidata entry for Mona Lisa<sup>8</sup>.



**Fig. 8.** Wikidata entry for Mona Lisa

<sup>6</sup> [www.wikidata.org](http://www.wikidata.org)

<sup>7</sup> [commons.wikimedia.org](http://commons.wikimedia.org)

<sup>8</sup> [www.wikidata.org/wiki/Q12418](http://www.wikidata.org/wiki/Q12418)

Every Wikidata entry has a unique identifier (here Q12418) and a label in various languages. All information is stored as key-value-pairs where keys are pre-defined properties, e.g. creator (P170), and values are either references to other Wikidata items, e.g., Leonardo da Vinci (Q762), or literals like a concrete birth date.

Wikidata is rich in information, both in terms of number of items and number of attributes for each item. The entry for Mona Lisa contains more than 200 attributes, including all links to Wikipedia pages about Mona Lisa.

From Wikidata, we extract the following entities:

- 110,000 artworks, e.g., Mona Lisa
- 21,000 motifs, e.g., mountains
- 16,000 artists, e.g., Leonardo da Vinci
- 4,800 locations, e.g., Louvre Museum
- 520 materials, e.g., oil paint
- 250 genres, e.g., portrait
- 220 movements, e.g., Renaissance

### 3.1 Data Model

For those entities, we selected attributes of interest. **Fig. 9** shows the data model of openArtBrowser, i.e., all entities, their attributes and associations as a UML class diagram.

The central entity is Artwork with subtypes Painting, Drawing and Sculpture. All entities have an id, a label and optionally a description. Different entities have different other optional attributes, e.g., height and width for an Artwork, date\_of\_birth and date\_of\_death for an Artist, latitude and longitude (lat, lon) for a Location, etc. The central entity Artwork is associated with the entities Material, Genre, Movement, Artist, Location, and Motif.

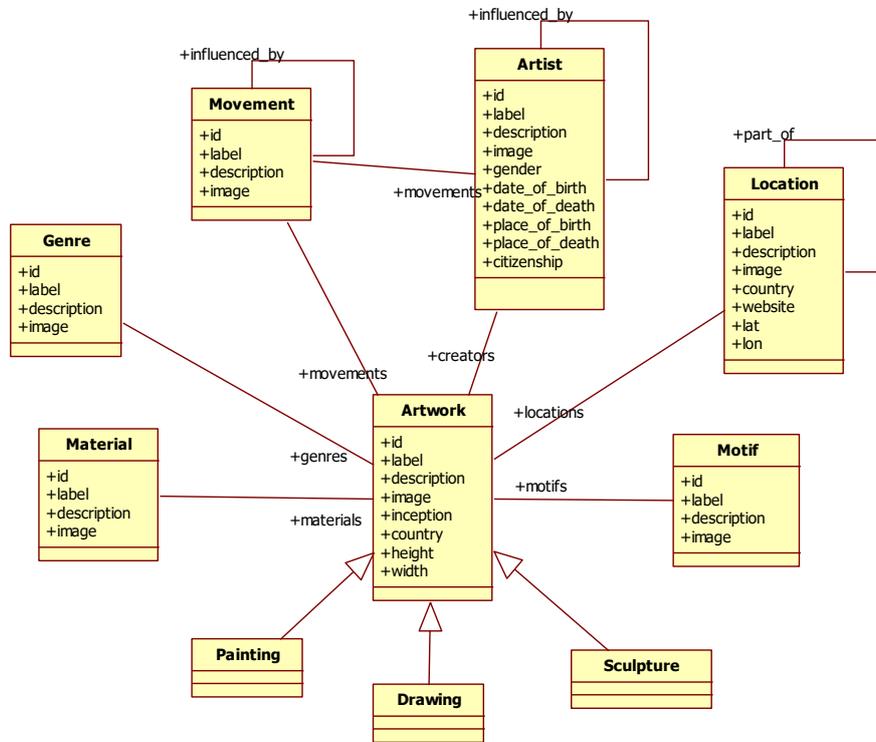


Fig. 9. Data model

### 3.2 Data Modeling Considerations

How did we select those entities and their attributes out of hundreds of entities and attributes provided in Wikidata? This was the result of a lengthy process inspecting the data provided. We used the following selection criteria.

- *Requirements*: Entities and attributes were selected which were deemed relevant for meeting the requirements of an art browser, i.e., information which is interesting to users. For example, information about artists and their backgrounds is certainly interesting. Also interesting is information about motifs depicted. Information we omitted were, e.g., the inventory number of artworks, the course of death of artists, the director / manager of museums, etc. Such attributes were not deemed relevant.
- *Quality*: We carefully checked the quality of data provided for certain attributes since the quality of crowdsourced data may vary considerably. For example, we observed the “instance of” and “subclass of” relationships of motifs. These attributes allow to model hierarchies of broader and narrower terms. This could potentially be useful. Users of openArtBrowser could, e.g., search for a motif “animal” and find artworks which are tagged with motifs “dog”, “cow”, “horse” etc.

(but not explicitly tagged “animal”).

However, we decided not to include this feature in openArtBrowser. The reason being that, due to a lack of agreed modelling guidelines, “instance of” and “subclass of” relationships are used inconsistently, nearly arbitrarily by the Wikidata community.

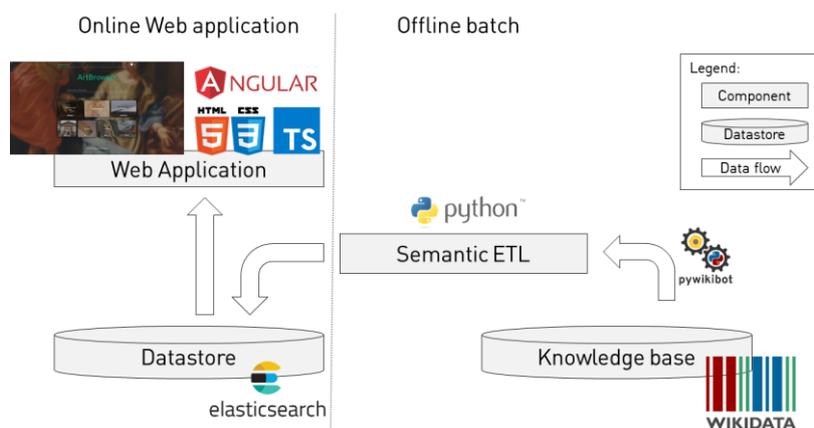
If we had blindly included hierarchical search according to the “instance of” and “subclass of” relationships, the search for motif “animal” would have also resulted in artworks that are tagged with motif “wife”. Why is this the case? This is because in Wikidata, the following relationship chain is modelled: wife is subclass of woman, is subclass of female, is subclass of Homo sapiens, is subclass of omnivore, is subclass of animal<sup>9</sup>.

*Quantity:* We also took into consideration, how frequently certain attributes are being tagged in Wikidata. For example, the Attribute “Iconclass notation” of artworks is relevant for expert users of openArtBrowser, since it indicates the iconography of artworks. The data quality is also good. Insofar, iconography would be a candidate for another facet. However, Iconclass notation is so rarely tagged so that it would be frustrating for a user to select this facet and then being able to navigate to only one or two other artworks.

It is worth noting that all three criteria, requirements, quality and quantity, may vary over time. Therefore, adapting the data model should be considered regularly.

## 4 Software Architecture

The software architecture of openArtBrowser consists of the online Web application and an offline batch. **Fig. 10** gives an overview.



**Fig. 10.** Software architecture of openArtBrowser

<sup>9</sup> Accessed 4/3/2019

The *online Web application* is designed as a two-layer architecture with the presentation and application logic being implemented in HTML<sup>10</sup> / CSS<sup>11</sup> / TypeScript<sup>12</sup> using the Angular<sup>13</sup> Web application framework. The *datastore* is implemented using the search engine Elasticsearch<sup>14</sup>. Queries to the datastore are executed very fast within a few milliseconds.

The *offline batch* implements a *semantic ETL (extract, transform, load) process*, which can be seen as the *curation process* of the arts data. In this process, relevant data is extracted from the *knowledge base* (Wikidata), is semantically enriched and transformed so that it can be loaded into the datastore (Elasticsearch). The offline batch process can be started regularly (e.g., weekly) and executes for more than 24 hours. This relatively poor performance is due to the response time of the Wikidata server. However, this does by no means affect the performance of the online Web application.

Semantic ETL is implemented with the Python programming language<sup>15</sup>. For *extracting* relevant data from Wikidata, the Python framework Pywikibot<sup>16</sup> is used. Extraction is based on the data model as depicted in **Fig. 9**.

*Data cleansing* is an important part of the extraction process. Since Wikidata attributes are not statically typed and Wikidata is crowdsourced, the quality of entries varies considerably. During the extraction process, a syntactic quality check is performed and data entries which do not conform to the expected data types are omitted. For example, references from artworks to artists which are no proper Wikidata links are omitted. The same applies for inception dates which are no integer numbers. Extracted data is stored as JSON files.

*Semantic enrichment* means adding value to the raw data. This includes computing custom ranks of artworks and all facets. The ranking criteria are described in Section 2.5. All ranks are normalized to values between 0 and 1 and get evenly distributed, so that the median always gets the rank 0.5. This enables comparing the ranks of artworks, artists, movements, genres, motifs, etc. Semantic enrichment also includes adding data from other sources, e.g., Wikipedia, Youtube, Iconclass, etc.

*Transformation* means storing the arts data in a format required by the datastore (Elasticsearch). In this case, this is a JSON format which reflects the data model as depicted in **Fig. 9**. Associations between entities are represented as arrays of Wikidata IDs.

Finally, *loading* is the step of updating the Elasticsearch index with newly extracted data. This is done using the Elasticsearch Update API which ensures continuous operation of the Web application.

---

<sup>10</sup> [www.w3.org/html](http://www.w3.org/html)

<sup>11</sup> [www.w3.org/Style/CSS](http://www.w3.org/Style/CSS)

<sup>12</sup> [www.typescriptlang.org](http://www.typescriptlang.org)

<sup>13</sup> [angular.io](http://angular.io)

<sup>14</sup> [www.elastic.co/de/products/elasticsearch](http://www.elastic.co/de/products/elasticsearch)

<sup>15</sup> [www.python.org](http://www.python.org)

<sup>16</sup> [doc.wikimedia.org/pywikibot](http://doc.wikimedia.org/pywikibot)

## 5 Discussion

We discuss openArtBrowser by comparing the implementation with the goals set out in the introduction.

1. *Learning with fun*: So far, no systematic user test has been performed and evaluated and insofar, there is no scientific proof yet that this goal has been reached. Instead, openArtBrowser has been tested in an adh-oc manner by various user and age groups. We observed users having fun discovering unexpected aspects in artworks and following tags, particularly motifs that were as yet unknown to them.
2. *Open Data*: All data displayed in openArtBrowser is open source from Wikidata and Wikimedia Commons.
3. *Serendipity*: There are various aspects of serendipity in openArtBrowser. The home page displays tiles with changing artworks, artists, movements, motifs etc. including their images. They invite users to follow interesting topics out of curiosity, without actively searching. In order to surprise also regular users over and over again, the home page and the tile contents change each time. All pages contain a gallery of artworks which attract attention. Those galleries are animated and slide from time to time (10 seconds). Additionally, artworks are shuffled on each use so that frequent users still can discover new aspects. The artwork page displays a gallery of related artworks where different relations are offered: motif, genre, movement, etc. Information about artworks are displayed as tags with hyperlinks. Clicking on those tags allows for discovering new aspects. Finally, also the semantic autosuggest feature fosters serendipity as various options for completing and refining the search terms are offered. However, the fact that search results are ranked and only the top-ranked entities are presented may lead, even with shuffling, to a filter bubble. This means that highly ranked entities get displayed regularly, but lowly ranked entities seldom or even never.
4. *Usability*: The application has been designed to be as simple as possible. Icons are used consistently in the semantic autosuggest, in tags, and in headlines of pages. Focus is on relevant information. E.g., on the artwork page, detail metadata is hidden when opening a page in order to avoid information overload. The response time is less than 1s for each page access, even when lots of information are displayed. Only when displaying images in full screen mode, high-resolution images are loaded which may take a little longer. Unser interaction follows common conventions, e.g., tags being clickable. Unusual interactions like comparing metadata of two artworks in the related artwork section are explained with a hint. An about page explains the use of the Web application in text form.
5. *Aesthetics*: We consider the Web application to be aesthetic and first users of openArtBrowser confirm this impression.
6. *Responsive Design*: openArtBrowser is responsive and can be used on various devices, including desktop computers, tablet computers, and smartphones.

So, we conclude that openArtBrowser meets the goals set out in Section 1.

The openArtBrowser implementation could also be used to implement a custom deployment with your own selection of artists and artworks. For this, the GitHub project would have to be forked and a filter would have to be implemented in the semantic ETL process.

Furthermore, the openArtBrowser concept and implementation could be used to implement semantic browsers for other application domains, e.g., movies, events, science, literature, history, politics, etc. For this, the GitHub project would have to be forked and the data model, the semantic ETL process, and the Web application would have to be adapted.

## 6 Conclusions and Future Work

We have presented openArtBrowser, a Web application for educating in visual art, fascinating users for paintings, drawings and sculptures. OpenArtBrowser is solely based on linked open data and its source code is open source.

OpenArtBrowser is actively being developed further. At the time of writing, the following features are being implemented:

1. *Multi-language support*: The first implementation of openArtBrowser was in English: dialog controls as well as metadata. We are currently implementing support for additional languages, namely German, French, Spanish, and Italian.
2. *More data sources*: Using semantic interlinking, openArtBrowser is enriched with additional data, e.g., Wikipedia abstracts. Care is being taken that this will not diminish the focus on relevant information and will not result in information overload.
3. *Multimedia*: YouTube videos about artworks, artists and artistic movements are integrated in openArtBrowser.
4. *Analytics*: User interactions are being logged in order to learn about the behavior of users and potentially improve user experience.

Future work will elaborate on this. In particular, Wikidata's identifier links to other knowledge graphs can be used to enrich data, e.g., WikiArt<sup>17</sup>, Europeana<sup>18</sup>, and Getty<sup>19</sup>. The filter bubble effect may be reduced, e.g., by regularly displaying some lowly ranked entities as well. Also, we could experiment with other ranking criteria like the Wikidata rank or usage statistics.

In addition to adding features to openArtBrowser, we intend to perform a thorough user evaluation and expect to learn insights for further improving the Web application.

Visit [openartbrowser.org](http://openartbrowser.org) and discover the fascinating world of visual arts!

---

<sup>17</sup> [www.wikiart.org](http://www.wikiart.org)

<sup>18</sup> [www.europeana.eu](http://www.europeana.eu)

<sup>19</sup> [www.getty.edu](http://www.getty.edu)

## References

1. Bernhard Humm, Timm Heuss: "Schlendern durch digitale Museen und Bibliotheken - Vom Umgang mit riesigen semantischen Daten" (in German). In Börteçin Ege, Bernhard Humm, Anatol Reibold (Editors): *Corporate Semantic Web*. Springer-Verlag, 2015. ISBN 978-3-642-54885-7.
2. Timm Heuss, Bernhard Humm, Tilman Deuschel, Torsten Fröhlich, Thomas Herth, Oliver Mitesser: *Semantically Guided, Situation-Aware Literature Research*. Workshop on User Interaction built on Library Linked data (UILLD 2013), Pre-conference to the 79th World Library and Information Conference, Singapore, 2013. In H.G. Cervone, L. G. Svensson (Eds): *"Linked Data and User Interaction"*, pp 66-84. Walter De Gruyter GmbH, Berlin / Boston, 2015. ISBN: 978-3-11-031692-6
3. Tilman Deuschel, Timm Heuss, Bernhard Humm: "The Digital Online Museum". Proceedings of the 4th International Workshop on Semantic Digital Archives (SDA 2014). London, UK, September 2014.
4. Tilman Deuschel, Christian Greppmeier, Bernhard Humm, Wolfgang Stille: "Semantically Faceted Navigation with Topic Pies". Proceedings of the 10th International Conference on Semantic Systems (SEMANTiCS 2014), Leipzig, Germany. ACM Press New York, USA, 2014. ISBN: 978-1-4503-2927-9, DOI: 10.1145/2660517.2660520.
5. Tilman Deuschel, Timm Heuss, Bernhard Humm, Torsten Fröhlich: "Finding without Searching - A Serendipity-based Approach for Digital Cultural Heritage". Proceedings International Conference on Digital Intelligence (DI 2014), Nantes, France, 2014.
6. Chantal Eschenfelder, Karsten Gresch, Torsten Fröhlich, Bernhard Humm, Thorsten Greiner, Peter Eierdanz, Frank Blumenberg: "The other way round: from semantic search to collaborative curation". Nordic Digital Excellence in Museums Conference (NODEM 2013), Stockholm, Sweden, Dec. 2013. Author, F.: Article title. *Journal* 2(5), 99–110 (2016).
7. Ulrich Beez, Bernhard G. Humm, Paul Walsh: "Semantic AutoSuggest for Electronic Health Records". In: Hamid R. Arabnia, Leonidas Deligiannidis, Quoc-Nam Tran (Eds): Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence. Las Vegas, Nevada, USA, 7-9 Decemeber 2015. IEEE Conference Publishing Services 2015. ISBN 978-1-4673-9795-7/15, DOI 10.1109/CSCI.2015.85