The recommendation of cultural heritage objects in Europeana Collections

Master thesis

Karl PINEAU

Master thesis director:
Pierre-Antoine CHAMPIN (University Lyon 1)

Europeana director:
Antoine ISAAC

Master of information architecture domain
Ecole Normale Supérieure de Lyon

École Normale Supérieure de Lyon
Academic year 2016-2017

Abstract

In this thesis, we present the results of an experiment run via a chatbot survey on Facebook Messenger which aimed to evaluate the preferences of the end-users of Europeana Collections about the recommendations of cultural heritage objects.
We highlight the ability of a small group of metadata (composed of dcType, dcSubject, dcCreator) to increase the number of clicks on a recommender system. We also highlight a probable ability of the users to be aware of their logic in the choice of recommender systems, and their trend to choose similarity (in place of serendipity) when there is no point of comparison. Finally, we point some advantages and disadvantages of surveys on chatbot.

Acknowledgements

This work would not have been possible without the support of many people.
First, thanks to my master thesis director Pierre-Antoine Champin who helped me and supported me all along the eight months of the thesis.
Then, I thank the Europeana Foundation which welcomed me for 6 months of internship in The Hague and more specifically the R&D team of the Foundation, with Antoine Isaac who took much time to supported me and review my work, and also Timothy Hill, Valentine Charles, Hugo Manguinhas and Nuno Freire. This work would definitely not have been possible without their support and feedback. I also thank David Haskiya, product and services director of the Europeana Foundation, for his comments and feedback. Moreover, I would like to thank all the European members of the Europeana Tech Committee for their advice, especially the team of Paul Clough at the university of Sheffield, Juliane Stiller of the Humboldt-Universität of Berlin.
I also thank my teachers of the master of information architecture domain of Ecole Normale Supérieure de Lyon and especially Jean-Michel Salaün for his advice and feedback during our thesis seminars. I don't forget my classmates of master thesis, for their reviews of my thesis. Also, I thank Emmanuelle Bermes, deputy director for services and networks at BnF and member of the Members Council of the Europeana Association, which consented to do an interview about the policy of data providers and the purpose of digital libraries.
Even if they possibly didn't know the purpose, I also thank the fifty people who kindly contributed to the experiment presented in this thesis.
And finally, thanks to my family for their support and encouragements.

Table of contents

1. Research problem

1.1. The information retrieval in the context of the development of the digital libraries

During the past two decades, we have witnessed the development of digital libraries in the cultural heritage domain. One of the first created in Europe was the French digital library Gallica[1] from the Bibliothèque Nationale de France, in 1990, and one of the biggest is Europeana[2]. These libraries gather millions of documents. They require reflection about their usages and the expectations of their users (casual users just as well as domain experts) about the retrieval of their documents. This is a challenge, especially in the context of Europeana, which gathers around 53 million items from around 25 countries and languages.

In this context, the recommender systems are investigated to facilitate information retrieval of documents and to help the user browsing in the library.
Indeed, the recommender systems are currently one of the most common functions on the web. The recommendation domain covers a large diversity of systems according to the product field. Social networks try to extend one's network with the "Do you know them" function (such as on LinkedIn[3]). Encyclopedias focus on serendipity (like the French version of Wikipedia[4]) and the question and answer sites call it "Linked or related topics" (like on Stack Overflow[5]), using term matching or semantic matching. Search engines try to find relevant similar queries (like the Google Knowledge card). If the technologies and the methods change, the common aim of these systems is to help the end-user browsing through a large amount of data and documents.

1.2. The recommendation of objects in Europeana Collections

The Europeana Foundation is a Dutch foundation mostly funded by the European Commission which hosts a European digital library portal, known as Europeana Collections. The Europeana Foundation also hosts many of other projects in plus of the digital library. Lots of Europeana involvements are not opened to end-users[6]. For instance, the Europeana Foundation provides a REST API service[7] for web developers in order to query the library.

[1] http://gallica.bnf.fr/
[2] http://europeana.eu
[3] https://www.linkedin.com
[4] https://fr.wikipedia.org/, e.g. in https://fr.wikipedia.org/wiki/Europeana, requires to login
[5] http://stackoverflow.com is an English popular board focused on new technologies
[6] We understand "end-users" here as users who "typically do not possess the technical understanding or skill of the product designers" such as in [https://en.wikipedia.org/wiki/End_user, 04/14/2017]. Here, they are the customers of the products based on Europeana Collections.
[7] http://labs.europeana.eu/api/introduction

The Europeana team also develops bespoke software like METIS[8], the data publication framework of Europeana Collections.
Last but not least, the Europeana Foundation is also a projects hub. It is a structure allowing organizations to connect and build projects. For instance, Europeana Sounds[9] is a Europeana project driven by national libraries (such as the British Library or the Bibliothèque nationale de France), sound research centers (such as CNRS laboratories), non-profit organizations (such as Europeana), universities and private companies.

The three main distinguishing characteristics of the digital library Europeana are its size (53 million documents), its multilingualism (there are end-users from all the European countries and documents are described in around 30 languages) and the heterogeneity of the objects (in Europeana Collections, there are famous paintings, as 3D models of doors, as, books from the eighteenth century, as photo of fashion show).

Currently, there is a recommender system available on Europeana Collections, on the page of the records [Fig. 1]. Its name is "Similar items" and it provides some Cultural Heritage Objects (CHOs) related to the one being displayed (named reference item). The aim of "Similar items" section is to provide a strict similarity. It means the current function will not create serendipity or diversity in its results. This function doesn't target any specific user.
In the Europeana context, little work has been run on this function by research teams related to the Europeana team. [Clough, Otegi, Agirre, and Hall, 2013] presents an experiment of contextualization of the recommendation, in the model of "people who viewed this item also viewed this item", run on Europeana objects in the context of the PATHS project. Also, the Europeana team has started a work (as described in [Hill, Charles, and Isaac, 2016]) on serendipity and expectations of the end-users. And advanced assessments in [Pineau, 2016e] of the function point to various issues, like the variation of the recommendations depending on the language of the interface. However, there is still no formal definition of similarity in Europeana nor as evaluation criteria. Moreover, there is no object curation in the current "Similar items" function (e.g. there is no selection of specific Europeana entities for the "Similar items" function). It means each entity of Europeana is eligible to this function, even if it doesn't have a thumbnail or a title (the two main metadata fields displayed in the "Similar items" section).

The recommendation of CHOs in Europeana is at the crossroad of several domains. Indeed, the recommendation depends on the recommender system domain which is a subdomain of search engines and information filtering system. But the CHOs context implies a modeling side as well as a digital humanity side. Finally, the end-user side refers to UX design approaches.

[8] https://github.com/europeana/metis-framework
[9] http://www.europeanasounds.eu/

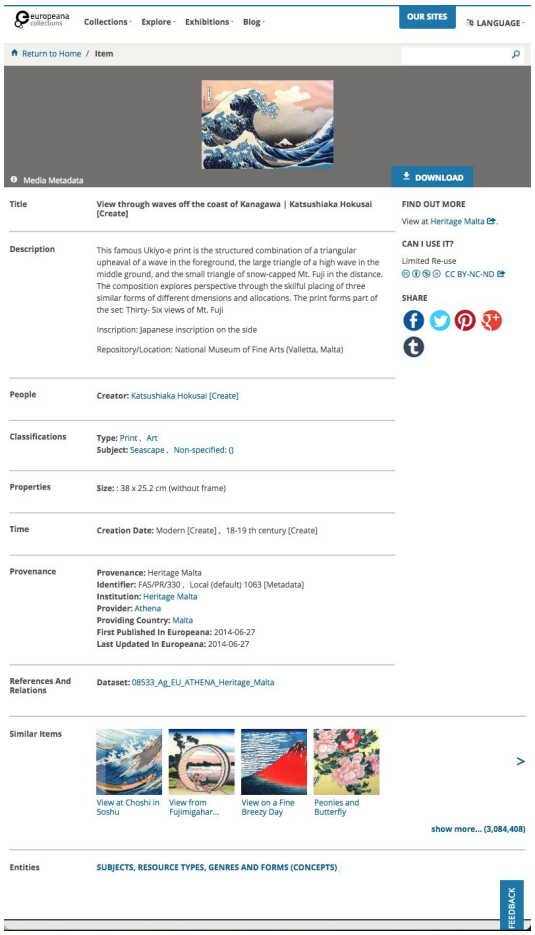


Fig. 1: The record page of Europeana Collections. The “Similar items” section is on the bottom part.

1.3. The difficulty of understanding the expectations of users about recommendations

Recommendation systems try to extend the user’s visibility on the products they recommend. For instance, when Stack Overflow presents its section “Linked or related topics”, it says to users “if you haven’t been able to find the answer about your issue on this topic, maybe you could find it on these ones”. In this case, Stack Overflow tries to provide a better answer to the user’s question. This question was previously asked to a search engine (Stack Overflow search engine or another search engine like Google). So this function is here to compensate wrong or partial answers from search engines, or to extend the first search query.

As for the other contexts previously presented, in the Europeana context, finding out what the user is looking for can be hard. Indeed, it depends on the user’s profile. We don’t know if the user is:

- Looking for a specific item: e.g. I am looking for Mona Lisa. The first search result with a strict similar label (search in French: “La Joconde”) is not the real Mona Lisa but an opera of this title. In this case, I could expect that the first similar item was the real Mona Lisa.
- Looking for a specific topic/collection: e.g. I am looking for furniture from the

eighteenth century. After browsing the first relevant item, I would like the recommender system to take into account the initial query to provide relevant items on my initial search, enriched by the current browsed result.

- Interested in related items/information about the displayed CHO: for instance, I am interested in capitals of columns and I would like to know more about the church of a browsed capital. Here, contextual entities would be really relevant to provide a new kind of information to the user.
- Interested in a catalogue raisonné. For instance, I am working on Ambroise Dubois (French painter in the court of François I^{er}) and I would expect that similar items to Ambroise Dubois' drawing would be other artworks from him.
- Interested in discovering the richness of Europeana (like the "random article" function of Wikipedia): for instance, I found a beautiful picture on Europeana via Google. This is the first time I come on Europeana and I am curious about Europeana objects. I would expect seeing other nice images, even if there are no relationships between them.
- Interested in other topics we don't know. This list is not exhaustive.

But Europeana doesn't have other information such as the profile of the user (there is no registration system for users in Europeana), and because of the three main characteristics of Europeana previously exposed (lots of objects, in many languages, from multiple domains), it is hard to understand what does a user look for, and to keep this information for future sessions. We can see it is really difficult to evaluate the interest of a user for a specific topic.

In this work, we try to better understand the expectations of users. In order to be able to identify a specific group of users, we chose to target only the users coming from social networks. Indeed, this is an important category of Europeana users. The team of Europeana defines in ['Sharing our validated learnings & embedding the validated way of working', 2016] the casual user of Europeana Collections as a social network user, looking for cultural content[10] to consume. This persona is someone young, creative, with a high level of education and is digital native. Targeting this category of users is important for Europeana because they can turn on into culture vultures (it means users who consume lots of cultural heritage content) and they are able to share and promote Europeana to their networks as defined in [Shen, 2014]. Plus, we define below that recommender systems are valuable only for some categories of users which the users from social networks are.
According to the Google Analytics account of the Europeana Foundation, social networks represent 3,66 % of the Europeana Collections frequentation[11]. If it represents a small

[10] It is important to specify that in the Europeana context, content is understood as a digital representation of a cultural heritage object. So content is most of the time the web resource of a CHO, using the notions of the Europeana Data Model. But in this thesis, we use the term "content" to define any group of resource, data or metadata constituting an information, e.g. an entity. We justify this choice by the common acceptation of the expression "content-based recommender system", which refers to our definition. When we will need to talk about Europeana content, we will use the expression "web resource".
[11] This measure has been done the 6 Mars 2017 for the period 04/06/2016-04/06/2017. 3.66% represent 118 227

quantity of sessions, the quality of those (we mean here the average duration - 02:52 - and the average ratio pages/session - 3,95) is higher than for the two main acquisition channels: organic search (duration: 02:11; pages/session: 3,04) and direct (duration: 02:34; pages/session: 3,48).
The Europeana team selects content (object from Europeana Collections, exhibition, blog post…) and promotes it on its social network accounts.
How can we improve the quality of the recommender system in Europeana Collections to invite the user coming from social networks to browse more Europeana Collections, to come back later and to become a promoter of Europeana Collections?

2. State of the art

The question of the recommendation of cultural heritage objects in the context of Europeana Collections implies three different parts for this state of the art.
First, a section will investigate the larger question of recommender systems. In order to provide a relevant state of the art for our purpose, we have decided to focus our work on a few methods and technologies of the recommender systems domain. Thus, we will not elaborate the question of the collaborative recommender systems[12], in order to focus on the question of the content-based recommender systems[13], which are the ones currently used in Europeana Collections.
Then, it seems important to specify the context of the Europeana Foundation and how it could be used by the recommender system to improve its results.

2.1. The recommender systems

The domain of the recommender systems is a research domain in information retrieval science. [Schafer, Konstan, and Riedl, 2002] define it as following: *"according to* [Resnick, Varian, 1997]*, 'in a typical recommender system people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients.' This definition includes three classes of systems. Suggestion systems provide a list of candidate items or recommendations. Estimation systems provide an estimate of user preference on specific items or predictions. Comment systems provide access to textual recommendations of members of a community."* According to this definition, we are going to focus our work on the first class of recommender systems, the suggestion systems, which are used by the recommender system of Europeana.

sessions.
[12] The collaborative recommender systems base their recommendations on recommendations made by other users.
[13] The content-based recommender systems base their recommendations on content similarity, and often similarity of profile of users.

2.1.1. Content-based recommender systems

Definition of a content-based recommender system

[Lops, Gemmis, and Semeraro, 2010] define content-based recommender systems as "*try*[ing] *to recommend items similar to those a given user has liked in the past. Indeed, the basic process performed by a content-based recommender consists in matching up the attributes of a user profile in which preferences and interests are stored, with the attributes of a content object (item), in order to recommend to the user new interesting items.*"

The content-based recommender systems include the first and the second classes of Resnick and Varian. The main difference between the two classes is that the first doesn't use user profiles when the second does use them.

Currently, Europeana hasn't built user profiles for recommendation or search. Then, we can classify its system in the first category. In Europeana, the system looks for similarities between metadata fields to define the recommendation. For instance, in the Europeana context, if the creator of the reference item is the Leonardo da Vinci, recommendations could be computed by similarity of creators and suggest other items from Leonardo da Vinci. In this case, recommendations will be items with Leonardo da Vinci as creator.

Of course, in order to define a very accurate recommendation, it will be necessary to use various metadata fields and various proximity ranges. Currently, in the Europeana context, recommendations are based on subjects, types, creators, title, and dataset fields. Various boosts are applied in order to adjust the weight of each field.

As previously said, a content-based recommender system can compute recommendations using a user profile. In this case, this profile would be the addition of the previous log of this user. For instance, a log could be compound of the items previously browsed. In this case, the recommendation based on the reference item would be refined by the user's profile. At this point, it is important to specify that the recommender system of Europeana doesn't use user profiles.

Keyword-based approach, the main part of recommendation in the Europeana system

According to [Lops et al., 2010]'s definition of content-based recommender systems, there are two main approaches to compute similarity of fields content. The first one is the keyword-based approach, which is currently the one mainly used in the Europeana recommender system. The second one is the semantic-based approach.

The keyword-based approach provides recommendation on keywords similarity, which is the simplest way to provide recommendation.

But according to [Lops et al., 2010], the main limit of this system is *"if the user, for instance likes 'French impressionism', keyword-based approaches will only find documents in which the words 'French' and 'impressionism' occur. Documents regarding Claude Monet or Renoir exhibitions will not appear in the set of recommendations, even though they are likely to be very relevant for that user."* In order to complete this highlighting of the limits of the

keyword-based approach, this is also a system which isn't valuable in a multilingual project as Europeana is. Indeed, if "French impressionism" results will be found, it won't be the case of "impressionnisme français" or "französischer Impressionismus". As one that should allow the system to make recommendations of English content for French people (and all other possible combinations of languages), the keyword-based approach is highly unsatisfactory in the Europeana context.

Semantic-based approach, the future of recommendation in the Europeana system

"More advanced representation strategies are needed in order to equip content-based recommender systems with 'semantic intelligence', which allows going beyond the syntactic evidence of user interests provided by keywords." [Lops et al., 2010]

Semantics-enabling of the metadata of Europeana Collections is a work that has been engaged by the Europeana team since few years, with the idea to enrich metadata provided by European organizations with IRI of controlled vocabularies or thesaurus such as the vocabulary Art & Architecture Thesaurus[14] (from the Getty Institute[15]) for the specific terminologies in art and architecture, the database geonames.org[16] for the geolocations, the vocabulary Union List of Artist Names[17] (ULAN, from the Getty Institute) for the information about artists, or now the contextual entities database of Europeana[18].

This system allows Europeana to provide search functions and recommender systems in multiple languages. Indeed, when Europeana fetches the IRI "http://vocab.getty.edu/ulan/500010879" of ULAN, the IRI returns the name of Leonardo da Vinci in many languages. It allows Europeana to render it in the language of the interface of the user. Plus, if Europeana stores in its database the IRI "http://vocab.getty.edu/ulan/500010879" in place of "Leonardo da Vinci", it makes easier to find every object from Leonardo da Vinci, according to the issue stated in the previous section.

Considering different recommender systems to better target the preferences of the users

The implementation of a meta-recommender system presented in [Schafer, Konstan, and Riedl, 2002] which could allow to define a recommender algorithm different (i.e., a different query in the search engine) according to the user's previous choices of similar items could be a response to the difficulty of understanding the preferences of users.

In this system, the "Similar items" block of the Europeana Collections portal should present various items from various recommender systems. For instance, if the "Similar items" block suggests four items to a painting of Leonardo da Vinci, the recommendations could be: the first recommended item from a recommender system based on the idea of providing a catalogue raisonné (i.e. another item from Leonardo da Vinci). The second recommended

[14] http://www.getty.edu/research/tools/vocabularies/aat/
[15] http://www.getty.edu/research/index.html
[16] http://www.geonames.org/
[17] http://www.getty.edu/research/tools/vocabularies/ulan/index.html
[18] http://labs.europeana.eu/api/entities-collection

item from a recommender system based on the country of the provider (let's say Italy). The third recommended item from a recommender system based on valuable items of the Europeana Publishing Framework (which gives value in particular to high-quality files for web resources and to free reuse rights). And the fourth could be a recommended item from a recommender system based on the matching of terms in the title. If the user choices a recommended item (that means, if she clicks on it), the next "Similar items" block will take into account the choice of the recommender system made and promote it in the next recommended items. E.g., if the user chose the item of the recommender system based on the country, in the next "Similar items" block, two items from this recommender system would be suggested.
Thus, the recommender system could better take into account the different usages of Europeana from the end-users and the difficulty to understand the preferences and the needs of the users that we previously demonstrate.

2.1.2. Main research problems in the recommender systems domain linked to the Europeana context

Cold start phenomenon

But the previous system based on meta-recommendations has an issue. Indeed, if there is no user profile on Europeana Collections, when the user leaves, we lost the information of what she is interested in. Plus, this system accepts that the first recommendations would not be deeply relevant in order to address the largest possibility of interests of the user. This is an issue we call the cold start phenomenon which is described in [Son, 2016]. The phenomenon deals with the issue of recommendation at the beginning of a user session, when we don't know a lot about the user.
Furthermore, the creation of user profile (a work in progress in Europeana) may not resolve this issue. Indeed, without user space (with functions like login, history of searches, etc.), it is hard to know when a user comes back and what were her previous sessions.
So, to provide good recommendations with user profile / meta-recommender system, it will be necessary to be able to manage with the issue of cold start phenomenon.

Data quality of Europeana content

Another important issue in the context of Europeana is the data quality. Indeed, issues of data quality provide outliers in the recommender system. As Europeana stores a lot of items, if the value of a metadata used by the recommender system is not really specific, it creates lots of noise in the results. For instance, a value "text" for the metadata "edm_dc_type" will fetch around 200.000 items in Europeana[19] and a value "monograph" will fetch around 6.200 items [20]. It is more difficult the rate the results from the general "text" value that for the specific "monograph" value.

[19] Number of results the 04-21-2017 : 193,234 | query
[20] Number of results the 04-21-2017 : 6,274 | query

Another issue is the multilinguality in the Europeana data (which is a part of data quality) where metadata are not in the same language. In the previous example, "monograph" (English value) will fetch around 6.200 items when "monographie" (French value) will fetch only 500 items[21]. This issue could be address by the semantics-enabling of the metadata of Europeana Collections.

2.1.3. Recommender systems in the domain of cultural heritage

Only few research projects have been run on the question of recommender systems in the domain of cultural heritage.

The CHIP project

The oldest one is the Cultural Heritage Information Presentation (CHIP) project[22] which aimed to provide recommendation of personalized museum tour, with Rijksmuseum data. But as the presentation highlights it, it focused on museum tour with a few hundred masterpieces, which is quite different in the approach than a database of 54 million documents.

The PATHS project

As previously presented, the Personalized Access to Cultural Heritage Spaces[23] (PATHS) project is probably the most reusable work done about recommender systems in cultural heritage domain for this current work. Indeed, the PATHS project collaborates with the Europeana foundation and works on its data and recommender systems. [Clough, Otegi, Agirre, and Hall, 2013] presents an experiment of contextualization of the recommendation, in the model of "people who viewed this item also viewed this item", run on Europeana objects in the context of the PATHS project. They highlight in this work the difficulty to provide recommendation (only 10.3% of the items have recommendations) due to the data sparseness.

2.1.4. Critic of the recommender system

The act of the recommendation can be easily considered as an action of curation. Indeed, the recommender system does choice into a set of objects the items presented to the user. Met in the context of an interview for this thesis, Emmanuelle Bermes, deputy director for services and networks at BnF, highlights that a curation action does not suit all categories of users. If it can be valuable for an end-user (we described below the users of Europeana) which doesn't know the collections of Europeana, E. Bermes highlights that a recommender system would not be such valuable for cultural heritage workers.
For example, we suggest above that some users of Europeana could be interested in a catalogue raisonné of an artist. In this context, the user is looking for an exhaustive list of objects, and not for a curated choice of objects.

[21] Number of results the 04-21-2017 : 490 | query
[22] http://chip.win.tue.nl/home.html
[23] http://paths-project.eu/

2.2. The specificities of Europeana Collections and the current work related to the recommender systems

2.2.1. Specificities of the Europeana data

In order to understand the specific problems of the recommendations in Europeana, we will describe here some of the unique characteristics of the Europeana data.

Data ingestion

Each record is provided by a European cultural heritage organization like a museum (for instance the Rijksmuseum Amsterdam[24]), an archive center (for instance the National Archives of Norway[25]) or a library (for instance the National Library of France[26]). "*More than three thousand institutions connect their digital collections to Europeana*" [Pekel, 2014]. The data ingestion process (which is described on its provider side in [Pekel, 2016]) sends data through various institutions processes, in a variable number of steps, and with variable end-results. These institutions are aggregators as defined in [de Hoog, 2014], which "*do the data harvesting from cultural heritage organizations. They also help to model the data, expand the network of organizations, help out with copyright queries and much more.*"
This situation with a lot of institutions induces a great heterogeneity (e.g. the language code normalization[27]) in the data encoding which makes difficult each Europeana data processing, including content recommendation.

Multilingualism of the records

[White Paper on Best Practices for Multilingual Access to Digital Libraries] specifies that, currently, Europeana provides records in around 25 languages. That means that there are metadata available in more than 25 languages in Europeana.
Each Europeana CHO metadata comes in at least one language. Most of the time, it is the language of the provider country. Of course, the record can be in more than one language. Depending on the language, the number of records can be really different. For instance, there are several million records in French (around six million records in November 2016) but only a few hundreds of thousands in Catalan (around three hundred and sixty thousand records in

[24] http://www.europeana.eu/portal/en/search?q=PROVIDER%3A%22Rijksmuseum%22
[25] http://www.europeana.eu/portal/en/search?f%5BDATA_PROVIDER%5D%5B%5D=The+National+Archives+of+Norway
[26] http://www.europeana.eu/portal/en/search?f%5BDATA_PROVIDER%5D%5B%5D=National+Library+of+France
[27] The language code was not normalized in the metadata of a CHO. For instance, Romanian items could be reference as "ro", "ron" or "rum" in the data. Europeana uses these codes to qualify the language of the metadata of a CHO or to qualify the language of a CHO. It is used to render a record as well as to compute recommendations. The described phenomena makes really difficult to find all the items described in a language if there is no normalization.

November 2016).

2.2.2. Architecture of the data of a semantic web leader

On a technical side, Europeana has a reputation of a semantic web leader in the cultural heritage domain. Indeed, Europeana provides the Europeana Data Model (EDM) [Charles, 2011]. This ontology aims to describes metadata of Cultural Heritage Object somewhere between the simple model of the Dublin Core[28] and the complex model of CIDOC-CRM[29], the two other main ontologies for cultural heritage, with a deep museum side for the CIDOC.
However, Europeana's "source of truth" is a Mongo database. Although these Mongo records are remediated into RDF triples, the resulting triplestore is a self-standing and independent service[30]; a similar remediation into Solr serves as the Europeana Collections backend.
To summarize and understand how the Europeana portal works, when a user asks for a record's page, the Europeana portal sends a request through the REST API to Solr index which contains and returns the information (the record) about the CHO. This index is frequently updated according to the state of the MongoDB database.
In a recommender system perspective, that makes it difficult to use semantic web-based approaches as the SPARQL endpoint is not deployed as core service and its sustainability isn't ensured.

2.2.3. The current works of the Europeana R&D team

Currently, the Europeana R&D team works on some important topics which could impact the computing of the similar items in Europeana in the next months.
First, Europeana is currently deploying a contextual entities database (the 'Entity Collection') which will contain entities like people, places, periods and will link them to the CHOs. This is a really important change because it will allow to provide contextual information to end-user. And [Murphy, 2016] shows that most of the searches are based on contextual entities and this contextual entities database will allow new functionalities like autocompletion. In a recommender system perspective, it will increase the semantics-enabling of the metadata of Europeana Collections and help to address the issues previously exposed.
Second, Europeana has just been elected as member of the IIIF[31] consortium and is one of the first organization supporting IIIF. In our context to provide better recommendations with better thumbnails, the IIIF technology should be an opportunity. Indeed, it can be used for the render of images stored on the server of a data provider which is sometimes hard.

[28] http://dublincore.org/
[29] http://cidoc-crm.org/
[30] http://sparql.europeana.eu/
[31] International Image Interoperability Framework is a framework for image interoperability which provides functions for the render of an image stored on a server.

2.2.4. The difficult question of identifying the users of Europeana

"For whom is your product intended?" seems to be one of the first asked questions by an investor to a team working on a new project. Understanding the usages of a product is indeed crucial to answer them the best way.
But this question seems to be complicated in the context of Europeana, considering the number of interests which are present at Europeana.

Some users can be stakeholders

As a European-funded project, Europeana has to respond to the aims of the European Commission. Those aims are synthesized in the visual report of a workshop about the users of Europeana, [Sharing Our Validated Learnings & Embedding the Validated Way of Working, 2016]. This report highlights that "*Europeana should benefit all* [European] *citizens*" and Europeana is "*funded as a public service*".
On the other hand, Europeana needs to persuade data providers to let Europeana ingest their data. Data providers have different aims than the European Commission. For instance, lots of them have expressed the need to show to their funding institution what they give to Europeana. Because of this, they need to find easily their data in Europeana Collections.

About the casual users of Europeana

Of course, the main part of the Europeana users is not composed of stakeholder users. But considering that there is no real service target user type, it is difficult to identify a specific type of users, or more precisely the aim of users.
In the visual report mentioned above, the Europeana team tries to identify casual users of Europeana. They define them as a "snacker-users", who use Europeana as a source of inspiration. These casual users use social media a lot, share information and are visual oriented users.

The culture vulture, the ideal user of Europeana

[Shen, 2014] defines another type of users of Europeana, the culture vulture, which is a user who consumes a lot of cultural content (movies, books, exhibitions, etc.).
"*The culture enthusiasts and professionals. They have a strong interest in cultural heritage and probably a good knowledge in a specific area(s). They are likely to work professionally with culture in one form or another, or to be a lifelong culture enthusiast, including researchers, students, professionals and interested laymen. While having a broad general interest a culture vulture has a special interest in, and knowledge of, one or a small number of specific topics, subjects, styles or genres.*
Culture vultures could come from any domains, such as an art student, a physical teacher, a musician, a journalist, a travel agent, a retired botanist, etc., usually with a good educational background. They have the needs of searching for resources about some topic(s) online and via other channels, to gain knowledge, expertise or inspiration." [Shen, 2014]

In our study, providing a good recommender system is important to ensure that culture vultures will find what they are looking for on Europeana and to help them better know the collections of Europeana.

3. Methodology

To answer our main research question which was "how to improve the quality of the recommender system in Europeana Collections to invite the user coming from social networks to browse more Europeana Collections, to come back later and to become a promoter of Europeana Collections?", we suggest, in an inductive method, to build an experimental protocol (that means to run a list of queries and to compare their results) in order to collect data about the recommended objects, according to the parameters of the recommender query.
We then suggest to collect user logs during experiments to measure the preferred choices of users between several algorithms. Based on that, we should be able to better understand the choices, the preferences and the understanding of users about the recommendations and the data of Europeana Collections.
In order to address our research problem, which is to better understand the user expectations about the recommender system of similar content of Europeana, we have followed the methodology described below.

3.1. Assessment of the current function recommender system of Europeana

A first part of our work has been devoted to a deep assessment of the current recommender system (the 'Similar items' card in Europeana Collections). [Pineau, 2016c] allowed us to understand the computation system of the algorithm. We had an overview of the weaknesses of the algorithm, such as outliers in the recommendations, e.g. the recommender system can suggest really different item but sharing one specific metadata, as the dataset; or on the contrary, the recommender system can recommend very similar items (like in Fig. 6). We also identified modifications of the recommendation according to the language of the interface (published in [Pineau, 2016e]), which was part of the issues. There are also saturation phenomena of the search engine score for some items, presented in [Pineau, 2016g], which is the result of certain data sparseness. Simultaneously, we have run a more general understanding of Europeana technical sides, as the semantic enrichment of the metadata as in [Pineau, 2016a], which is taken into account for the computing of the recommendations. All this work highlights the necessity of correcting few issues (such as the language of the interface) and the necessity of trying to have recommendations more understandable that actually.
For this first part of our work, we compared similar items of different records in different user interface language with the current algorithm, in order to test it, to find its shortcomings.

We published in [Pineau, 2016b] the result of the experiments where we tested the recommendation of the algorithm on datasets, first in an empirical way (e.g. do the images seem similar?), then in a way more structured (e.g. do the result scores form a long tail phenomena?)

3.2. The possible modifications of the search query of the recommender system of Europeana

This evaluation done, it had been necessary to find a way to improve the results of the algorithm. We agreed with the Europeana R&D team that it will be more effective working in an iterative way than trying to produce directly the best experiments. Indeed, it is difficult to quickly set up an experiment because of the quality of Europeana data.
Therefore, we produced an empiric list, based on our own knowledge of cultural heritage domain, but strengthened by criteria such as the Europeana strategy on data quality. The deliverable is a list of proposals of search query modification for the recommender system, published in [Pineau, 2016d]. For instance, one of the observations on the current algorithm was the prominence of the dataset information (that means the original group of items provided by an institution to Europeana) in the computing of the recommendations. Then, a proposal was to remove the dataset criteria of the search query in order to create more diversity in the results. Another proposal, based on the Europeana policy will to improve the data quality, was to force the algorithm weighting related items according to their quality, so that higher-quality items are preferred over lower-quality items. This quality has been defined in the Europeana Publishing Framework[32] and takes into account criteria such as the thumbnail availability or the reusable rights.
The following step has been to assess these proposals. In order to, we have run experiments where we computed (with the dataset previously presented) the similar items for each proposal of search query. This step resulted in a list for each proposal containing the top five recommended items for each item of our dataset, and a list for each proposal containing the search score for the first, tenth, hundredth and thousandth recommendations for each item of our dataset.
With this result, we were able to understand two different aspects of the query modification: is there a modification of the top results and is there a modification of score evolution? The final aim was to select a relevant set of search queries which are different from the current one in order to run an experiment about the perception and the preferences of the users about the search queries and the recommender system of Europeana.

[32] http://pro.europeana.eu/publication/publishing-framework

4. End-user evaluation of cultural heritage objects recommendation on Europeana Collections

4.1. Context of the experiment

This experiment is a part of the global aim of the improvement of the recommender system of Europeana Collections and especially of the improvement of the relevance of its results.

4.2. Aims

4.2.1. Users' preferences measure

The first aim of the experiment is to provide metrics on the preference of end-users of digital libraries in terms of content recommendation.
Thus, the experiment must demonstrate what is the receptivity of users to various recommendation queries - and with which level of information needed - which use different methods of recommendation.
So, we don’t want to measure the similarity between two items, we want to measure the user interest for this recommendation.

4.2.2. Gamification and low user effort

We want to enroll the highest possible number of users in this experiment. To make this possible we seek to follow two main principles: gamification and low user effort required from the user.
Regarding gamification, the main aim is to offer a playful experience to encourage the user to stay as long as possible. Here we use a chatbot on Facebook Messenger, and by renewing as much as possible the proposed content.
However, we don’t want that users feel weary of participating in this experiment. So we are not going to propose them a too complicated task. This means especially we should limit the range of user actions (e.g. clicks) that trigger new actions.

4.3. Research questions

Our experiment seeks to answer three main research questions and two secondary ones.

4.3.1. Question 1: Is it possible to identify a logic in a user's selection of recommendations?

This question aims to understand if the user choices in term of item browsing are motivated by the research of specific information or at least by the desire to look deeper into a

knowledge domain. Is the user interested in texts, or only items linked to Leonardo da Vinci, or does she look only for items of the 16th century?

To answer this question, I suggest in the first part of this experiment a user path in which items will be recommended according to a reference item. Each recommendation is provided by a different recommender system. The differences between the queries are the used metadata (see below).
My aim is to measure for one user the consistency with regard to the recommender system choice during the browsing, from one suggested items to another.

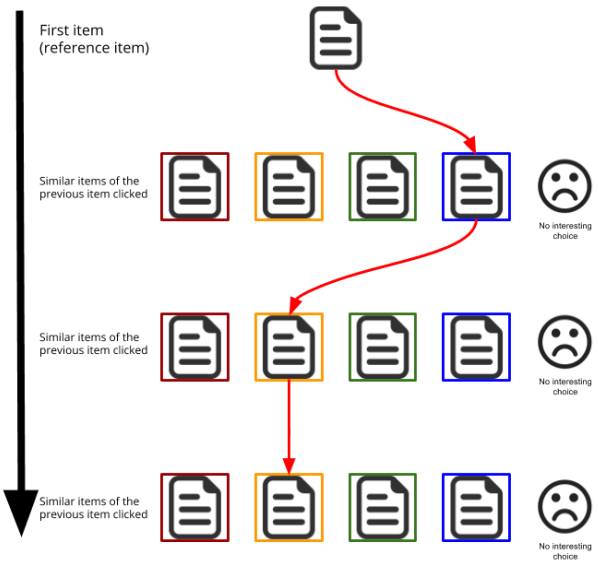


Fig. 2: Research question 1 - Second experiment - Is that possible to identify a logic in the data graph browsing of a user?

We ask the user to select the most interesting item in her opinion, between the presented. Each color represents one algorithm which could be:

- Wine color algorithm: weighting the creator field
- Orange color algorithm: weighting the Europeana Publishing Framework value
- Green color algorithm: weighting the date field
- Blue color algorithm: weighting the provider country field
- No interesting choice: the user can't find anything interesting in her opinion

Each similar item is the top-one of the corresponding algorithm. For each item, we present to the user the thumbnail, the title and a short description (if available) of the item. The user doesn't know that each item is generated by a different algorithm. Here, we can see user prioritizes the orange algorithm and the blue one.

To complete the answer to this first research question, I want to run a second experiment. The aim of this one will be to evaluate - outside of the context of browsing - if a user prioritizes the recommendations the same way than during the browsing. More generally, I want to know if the user has an intuitive ranking of the recommender systems, based on the rates given to their recommendations. Also, I want to know if between the recommender systems, one will emerge as the users' preferred one.

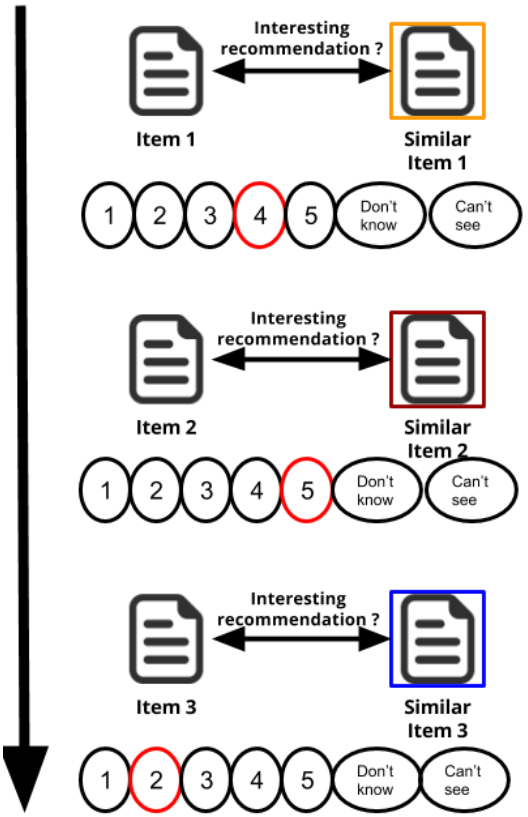


Fig. 3: Research question 1 - Second experiment - Is that possible to identify a logic in the data graph browsing of a user?

We ask the user to rate the recommendation in her opinion, between the presented. Each color represents one algorithm which could be:

- Wine color algorithm: weighting the creator field
- Orange color algorithm: weighting the Europeana Publishing Framework value
- Blue color algorithm: weighting the provider country field
- The "Don't know" button allows the user to express an unavailability to rate
- The "Can't" see button is used to tag objects and recommendations without image or title

Each similar item is the top-one of the corresponding algorithm. For each item, we present to the user the thumbnail, the title and a short description (if available) of the item. The user doesn't know that each item is generated by a different algorithm. Here, we could say that for this user, the wine color algorithm is more valuable than the blue one.

The result of this experiment should take into account the previous experiment about the Research question 1 to highlight favorite algorithms.

4.3.2. Question 2: Are users aware of their logic?

This second question aim to understand first if users are conscious of their choices, but also if the recommender system can orient them by giving context about the recommendation. By context, I mean here an information given to the user spelling out the recommendation, on the model of "this item is suggested because it has the same creator as the previous item".

To answer this question, I suggest completing the two experiments previously shown by two other versions, taking up the same principle but with context in recommendation.

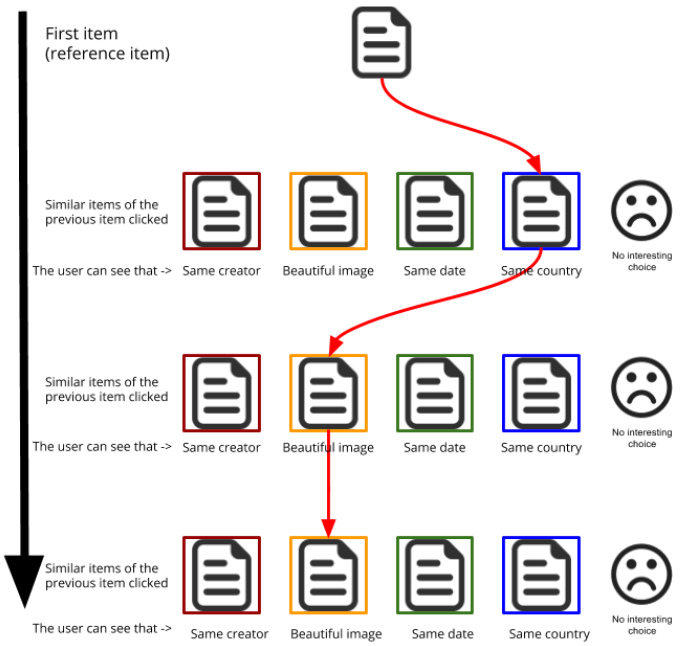


Fig. 4: Research question 2 - First experiment - Are users aware of their logic?

We ask the users to select the most interesting item in their opinion, between the presented. Each color represents one algorithm which could be:

- Wine color algorithm: boosts the creator field
- Orange color algorithm: weighting the Europeana Publishing Framework value
- Green color algorithm: weighting the date field
- Blue color algorithm: weighting the provider country field
- No interesting choice: the user can't find anything interesting in her opinion

Each similar item is the top-one of the corresponding algorithm. For each item, we present to the users the thumbnail, the title and a short description (if available) of the item. **We give information to the users about the similarity of the two items. That means, the users can see the shared metadata.** Here, we can see user prioritizes the orange algorithm and the blue one.

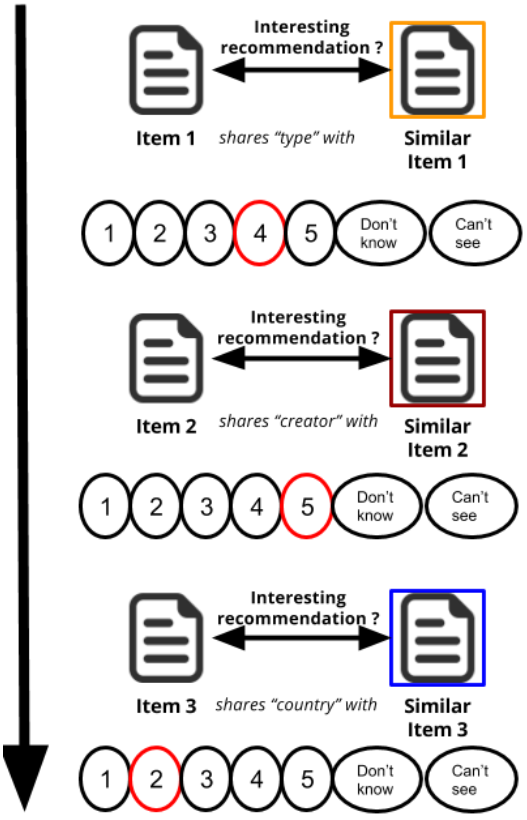


Fig. 5: Research question 2 - Second experiment - Are users aware of their logic?

We ask the users to rate the recommendation in their opinion, between the presented. Each color represents one algorithm which could be:

- Wine color algorithm: weighting the creator field
- Orange color algorithm: weighting the Europeana Publishing Framework value
- Blue color algorithm: weighting the provider country field

Each similar item is the top-one of the corresponding algorithm. For each item, we present to the users the thumbnail, the title and a short description (if available) of the item. **We give information to the users about the**

similarity of the two items. That means, the users can see the shared metadata. Here, we could say that for this user, the wine color algorithm is more valuable than the blue one.
The result of this experiment should take into account the previous experiment about the Research question 1 to highlight favorite algorithms.

4.3.3. Question 3: What are the categories of metadata that users "endorse" when they browse the recommendations?

This question aims to understand more specifically which categories of metadata are targeted by the end-users of digital libraries during their browsing. For instance, are they more interested in categories of metadata related to the dates, or to the agents.
The previous experiments will allow me to answer this question by mining the produced data.
The answer of this question can help in a further time to drive the current recommender system on popular properties.

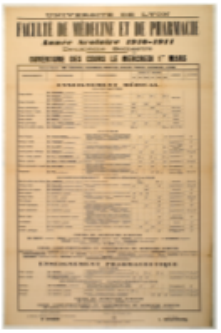
4.3.4. Question 4: (secondary question) Do the users consider the similarity between two items as a strict similarity or as a serendipity search?

The current recommender system for computing similar items in Europeana Collections has been designing to suggest (nearly) strictly similar items.
For instance, here an item (left) and its top-suggested item (right). We can see the title is similar at 80%, the metadata at 90% and the two images are close, visually speaking.

[Emploi du temps] Année scolaire 1908-1909. Premier semestre

typographie, cachet : "bibliothèque de la ville de Lyon", Université de Lyon ; ouverture des cours le mardi 3 novembre

Reference Item



[Emploi du temps] Année scolaire 1910-1911. Deuxième semestre.

typographie, cachet : "bibliothèque de la ville de Lyon", Université de Lyon ; ouverture des cours le mercredi 1er mars

Suggested Item

Fig. 6: Example of two objects with a high level of similarity

In this context, does a user consider this recommendation as a more interesting recommendation than a less similar item but giving more room to serendipity?
To answer this question, I am going to measure the similarity between metadata fields of items to compute a level or percentage of similarity between items in the recommendations used for the experiment. Then, it should be possible to compute an average of the

level/percentage of similarity of the similar items chosen by the users, which I will compare with the general level/percentage of similarity of the items composing the recommendations.

4.3.5. Question 5: (secondary question) Will a Facebook Messenger user be sensitive to a scientific evaluation action?

Chatbot usage is increasingly common on social media in order to propose new services to users to allow existing services to vary their access points in the way of a cross-channel approach.
In this context, chatbots could be an interesting medium for scientific research to help conduct experiments. We choose to run our experiment on Facebook Messenger, Facebook's instant messaging, using a chatbot.
One of the questions raised by this experiment is the readiness of Facebook Messenger users to participate a scientific evaluation action. This question will be evaluated with two parameters: people engaged compared to people reached[33], and the time spent / number of clicks done by each user.
This question also asks the problem of bias induced in the scientific experiment with a non-controlled environment as Facebook is. It is also necessary to consider Facebook constraints (see below).

4.4. Research method

4.4.1. Questions asked to users

I have presented in the previous section the experiment that I am going to run to answer the research questions.
We don’t want the user to judge if two items are similar. We are interested in knowing why users are interested in a specific type of recommendation (between two items).
Thus, the questions addressed will be the following:

- Experiment 1: “You are now browsing this object. In your opinion, which item in the following list seems the most interesting to continue your browsing?”
- Experiment 1 with context: “You are now browsing this object. In your opinion which item in the following list (where the first item shares the creator of the record, the second has a high-quality image, the third shares the type) seems the most interesting to continue your browsing?”
- Experiment 2: “Based on this object, we suggest you the following one. Can you rate this recommendation between 1 and 5? 1 means that you do not think this is an interesting recommendation, 5 means that you think this is a very interesting recommendation.”

[33] In the Facebook language, a "reached" person is someone who saw the message inviting her to contribute to this evaluation.

- Experiment 2 with context: “Based on this object, we suggest you the following one sharing the same creator. Can you rate this recommendation between 1 and 5? 1 means that you do not think this is an interesting recommendation, 5 means that you think this is a very interesting recommendation.”

4.4.2. Sequence of questions

One of the advantages of the chatbot (see below for details) is the ability to do a non-closed experimentation. Indeed, we can generate a new question to each user each time she answers one. Of course the user can still stop the experiment when she wants, with the ability to come back later and continue.
To structure the user path during the experimentation, I define “sessions”. One session is a list of questions from one of the four experiments (two experiments plus their 'contextualized' version).
For the first experiment (browsing of recommendations), a session size corresponds to the number of recommendations starting from the reference item. That means that for a reference item, there are 6 similar items. One is picked by the user and becomes the new reference item. Then we suggest to the user 6 other similar items for the new reference item and we continue until we don’t have any similar items to show. Considering the structure of the recommendations, this point should come after 4 suggestions. Indeed, I define below that similar items are generated for each reference item on four levels of depth.
For the second experiment (rating of the similarity of a couple of items), a session size is 5 questions. The number of five has been chosen to be nearly similar to the size of a session in the first experiment, and to not bore the user.

4.4.3. Computing recommendations

Initial dataset

To insure the system permanence during the experiment, the chatbot doesn’t require the REST API of Europeana Collections. We developed a specific API for this. This API fetches a segment of the metadata graph of Europeana Collections.
Ten items are used as base of this segment. To put this experiment in a realistic situation of recommendation for casual, I selected the last 10 CHOs shared on the Facebook page of the Europeana Foundation on Monday 13 February. These are the following:

1	http://www.europeana.eu/portal/en/record/2032004/2778.html
2	http://www.europeana.eu/portal/en/record/2021664/search_identifier_umg_items_2d688f4872f6db25b852ad5f6f35663b.html
3	http://www.europeana.eu/portal/en/record/9200365/BibliographicResource_1000055664026.html
4	http://www.europeana.eu/portal/en/record/2021641/publiek_detail_aspx_xmldescid_55194153.html

5	http://www.europeana.eu/portal/en/record/92062/BibliographicResource_1000126129416.html
6	http://www.europeana.eu/portal/en/record/15801/eDipRouteurBML_eDipRouteurBML_aspx_Application_AFFL_26Action_RechercherDirectement_NUID___53__AFFL_3BAfficherVueSurEnregistrement_Vue_Fiche_Principal_3BAfficherFrameset.html
7	http://www.europeana.eu/portal/en/record/2059209/data_sounds_C0023X0047XX_0600.html
8	http://www.europeana.eu/portal/en/record/9200365/BibliographicResource_2000081578901.html
9	http://www.europeana.eu/portal/en/record/9200365/BibliographicResource_2000081569228.html
10	http://www.europeana.eu/portal/de/record/08629/0105.html

Fig 7. Table of the items of reference for the experiment

Tested recommender systems

I have made the choice to test six different recommender systems based on queries sent against the Europeana REST API. It is necessary to state here that because the queries are done through the REST API of Europeana (in order to collect the items stored in the back-end of our API), we are not able to use all the Solr functions available for the "More Like This" function[34] of the API, especially the boost values which allows to prioritize some properties. Details on each algorithm are given in the appendix 2 of this thesis.

1. Recommender system derived from the current recommender system

A recommender system derived from the current recommender system of Europeana will be proposed as a baseline. Thus, it will be possible to evaluate other recommender systems against the baseline and to measure the possible effects of the others in a possible implementation on the portal.
The current recommender system works according to the following pattern:
Compilation of the parameters dcType, dcSubject, dcCreator, title, data_provider with a Boolean operator OR and a strict matching of values (restricted with quotation marks).

2. Recommender system based on a large set of metadata

The aim is to propose a recommender system which doesn't focus on specific metadata but takes all the descriptive metadata of an object into account. To not fall into the trap of a large query which will not prioritize its results[35], I suggest to open the query to the all the descriptive metadata and to gather them with a Boolean operator OR. But when there are several values for one metadata field (e.g. several creators, so several values inside the

34 https://github.com/europeana/europeana-blacklight/blob/develop/app/models/europeana/blacklight/document/more_like_this.rb#L18-L25

35 https://docs.google.com/document/d/1SvL1Px_xzzU9QdhZRMsR2YWPEjABMZ-FREQL3x1T7fM/edit?usp=sharing

metadata dcCreator), I combine them with a Boolean operator AND. It will force the query to find - for instance - CHOs with creator A AND creator B, OR type C AND type D.

3. Recommender system with selection of metadata

This recommender system aims to target a selection of metadata of a CHO: dcCreator, dcContributor, dcType, dcSubject, dctermsProvenance, dctermsSpatial. The aim is to bring curation in the content selection, in order to consider Europeana's collection heterogeneity.

4. Recommender system with chronological metadata

Under the same pattern than the previous recommender system, this one aims to chronologically gather content. Here again, the idea is to bring curation in the content selection.

5. Recommender system based on the Europeana Publishing Framework

This recommender system aims to recommend contents considered by the Europeana Publishing Framework[36] as highly valuable. To define this quality, the EPF takes into account the quality (size and resolution) of the web resource (image, text, sound, video, 3D object) and the ability to freely share the content.
This recommender system seeks to remove the CHOs with "poor" content to propose to the Europeana end-users an added value in the way of serendipity. It also invites data providers to share high-quality contents.

6. Recommender system with random item

Finally, as a basis for comparison, it seems important to put in the recommended items a random item to evaluate the added value of the recommendation. The aim of this item is, first to assess the value of the recommendation (if the users click on items from this algorithm, it probably means they can't find any interesting or valuable recommendations with the other algorithms), and second - here we make the assumption that the recommendation is valuable - the be a control item in the experiment.

Recommendations computing

For each of the ten initial items presented above, I compute the six similar items for the six recommender systems presented. I repeat the operation for each given item on four levels of deepness.
At the end, we get: $(10x6)^4$ items, i.e. theoretically a network of 12.960 items. Note that this is a theoretical maximum: there will probably be items, which will be recommended several times. In which case, they are going to be linked but not added once again. Also, a recommendation query may provide no result.

[36] http://pro.europeana.eu/publication/publishing-framework

4.4.4. Experiment environment

Using a chatbot for the survey

As previously said, we want to run this experiment on Facebook and Facebook Messenger services. This approach has pros and cons:
Pro: we are here in a pervasive approach in term of information architecture domain. The way to provide Europeana data on Facebook and to fetch the experiment data on another API is completely in line with a trans channel approach as defined by Andrea Resmini and Luca Rosati[37]. Indeed, the user has the possibility to choose her platform and Europeana doesn't ask her to break with her digital usages to contribute to this experiment. Plus, we can reach many users, as there are 306 million Europeana citizens inscribed on this social medium[38]. The Europeana Facebook page has around 90.000 "Like" mentions. If only 1% of these users would contribute to this experiment, it would be a success.
Pro: portability. The experiment can be run on Facebook, Facebook Messenger, but also their mobile application, on every mobile OS.
Pro: the possibility to send reminder message to a user without asking her email address.
Pro: the usage of Facebook for this experiment seems in agreement with the definition given by the Europeana team of a casual user. This is a young person, using social media, looking for inspiration or cultural content. In this context, the user is solicited by the chatbot for simple actions (a click), which can encourage the discovery of new contents, in a long period but without the need for regular actions.
Con: Facebook constraints the display of data. It is impossible to render data as we want. For instance, when we render an image gallery on Facebook Messenger, it is highly probable that Facebook will crop the images.

Recruitment of users

The aim is to recruit the Facebook users by messages on the main Facebook "places" in link with Europeana (the page of the foundation) and of digital humanities (like the French group MuzeoNum[39] which counts 2.500 members)

Gathering result data

The data generated on the chatbot during the experiment are sent to an API which mines it. In this way, it is possible - thanks to the chatbot - to follow a user (anonymously), from sessions to sessions, even if she comes back two weeks after a first session or she doesn't allow cookies on her web browser. It is also possible to collect data generated by the experiment: clicks, typed texts, dates, logs, failures. It is important to specify that personal

[37] Resmini and Rosati, *Pervasive Information Architecture - 1st Edition* <https://www.elsevier.com/books/pervasive-information-architecture/resmini/978-0-12-382094-5> [accessed 18 February 2017]
[38] https://zephoria.com/top-15-valuable-facebook-statistics/
[39] https://www.facebook.com/groups/muzeonum/

data of the user is partially available (name, first name, picture, sex, locale language and timezone are) but not collected (except the locale language and the timezone) in this experiment. Then, it is possible to use this collected data (structured in: user > sessions > questions > answers).

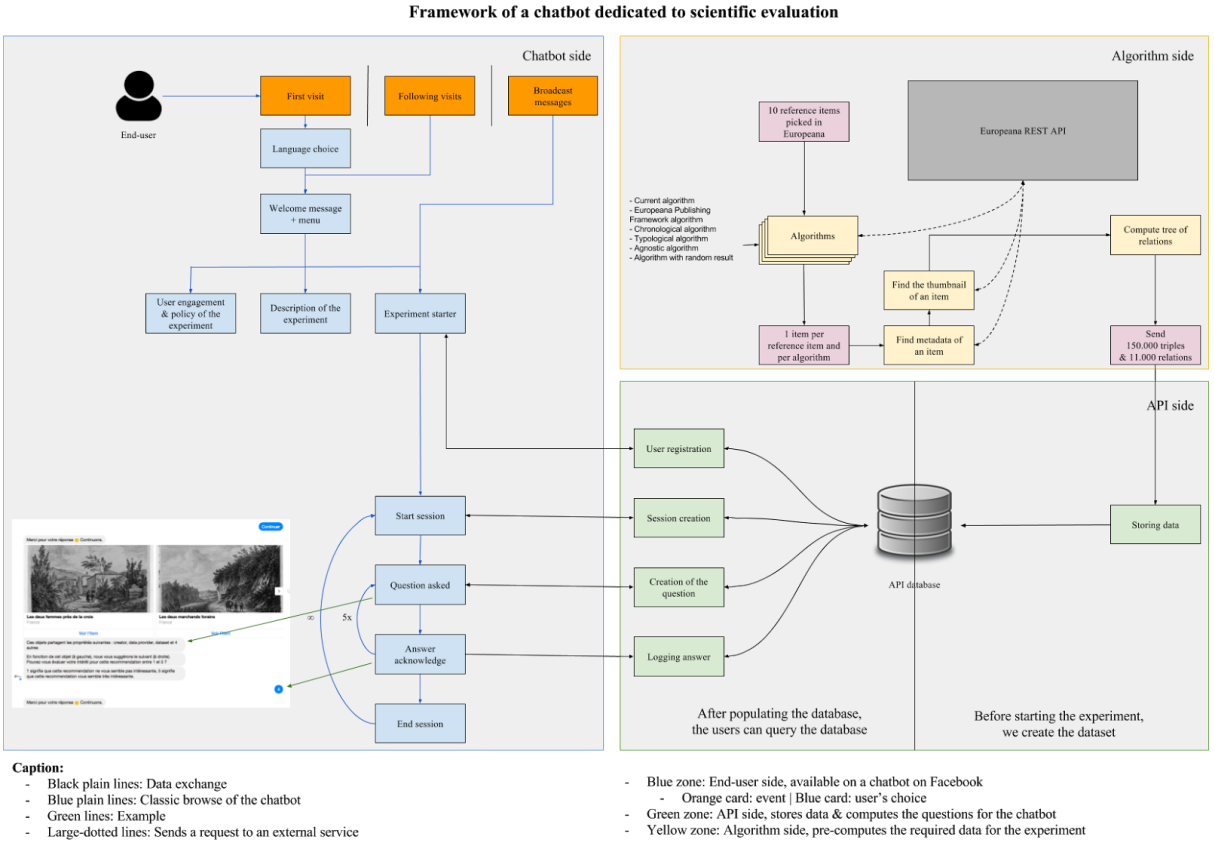


Fig. 8: schema of the process of the chatbot for the evaluation of the preferences of users

5. Results of the experiment

5.1. Conduct of the experiment

5.1.1. Announcements of the experiment

The experiment has been run from 27 March 2017 to 10 April 2017 with a chatbot on Facebook Messenger. The experiment has been announced on various media and supports on Facebook and other social media.
Here, a table of the different announcements done during the experiment:

Date	Media	Group/Page on the	Reachable people	Interaction with

		media		people
03-27-2017	Facebook	Ecole du Louvre	around 2.800	2 likes
03-27-2017	Facebook	MuzeoNum	around 2.850	5 likes
03-30-2017	Facebook	Europeana.eu	around 90.000	5 likes
04-04-2017	Facebook	Europeana.eu	around 90.000	7 likes
04-08-2017	Facebook	Europeana.eu	around 90.000	6 likes
03-27-2017	Facebook	Clichesart	around 850	8 likes
04-09-2017	Facebook	Clichesart	around 850	0
03-30-2017	Facebook	Europeana AllezCulture	around 670	1 like
04-03-2017	Twitter	@gsergiuc	around 60	1 like, 4 RT
03-30-2017	Twitter	@thillzilla	around 113	3 likes, 3 RT
03-27-2017	Twitter	@KarlPineau	around 160	2 RT, 1 like, 14 clicks,
03-30-2017	Twitter	@KarlPineau	around 160	4 RT, 6 likes, 11 clicks
04-03-2017	Twitter	@KarlPineau	around 160	2 RT, 1 like, 1 click
03-30-2017	Europeana Slack	General channel	around 70	0

Plus, to engage users to contribute regularly, the chatbot sent them a daily reminder. This daily reminder could be turned off if the user wanted to.

5.1.2. Translation of the experiment

The first days of the experiment have shown that the main part of the users was French users. To facilitate their usage of the chatbot, we have translated it into French two days after the start of the experiment.

Then, the chatbot was available in English and in French. The users had the possibility to select their favorite language at the start of the experiment.

All the content was translated, except the name of the metadata in the contextualized version of the experiment. But considering the English name of the metadata, we can assume that it was not complicated to understand it, most of them being transparent (e.g. “creator” for “créateur”).

5.1.3. Identified biases

Contextualized version of the experiment

Some feedback from users have highlighted the fact that they didn't see at the beginning of the experiment the information providing context on the results. That means they didn't see the sentence listing the shared metadata between two items, which was supposed to help them to rate the recommendation.
We don't know how many users didn't see this information, so the conclusions of this experiment related to the contextualization should be taken with precaution.

Variations in the engagement of users

The results' details show big differences in the engagement of the users. Some of them did only one or two sessions of experimentation when others did tens. More specifically, 3 users did around the third of the total amount of sessions. So, their opinion is over represented in the result of this experiment.

Quality of the thumbnails and the titles

It is important to specify that many recommended items didn't have a thumbnail or title. This issue could come from the API developed for the experiment which was unable to fetch all the metadata, because it didn't use the correct metadata to fetch a title. It could also come from the servers of the data providers which are not always query-able or because these objects don't have thumbnail. In this case, no thumbnail could be provided. This is not a minor issue in the experiment, especially for the thumbnails. Indeed, the users only have this information to do their choice. It affects probably between a quarter and a third of the items.

Rendering of the thumbnails

Considering that the experiment has been run on a Facebook Messenger chatbot, the thumbnails of the items have been cropped by Facebook Messenger. Sometimes, less than half of the image was displayed to the user. This creates a bias when a user can't compare two (or more) full images.

5.2. Presentation of the data

In the following results, to facilitate the reading of the graphs, we simplify the names of the recommender systems used:

- "Default" matches with the recommender system derived from the current recommender system
- "Typological" matches with the recommender system with selection of metadata
- "Agnostic" matches with the recommender system based on a large set of metadata
- "Chronological" matches with the recommender system with chronological metadata

- “EPF” matches with the recommender system based on the Europeana Publishing Framework
- “Random” matches with the recommender system with random item

We also used “browse’ sessions” to name the first experiment where the user chooses the most relevant item in her opinion between 6 choices, and “single’ sessions” to name the second experiment where the user rates one recommendation.

The row data of the experiment is available at the following address.

5.2.1. Users

The experiment has been completed by 53 users. They are distributed in the following way.

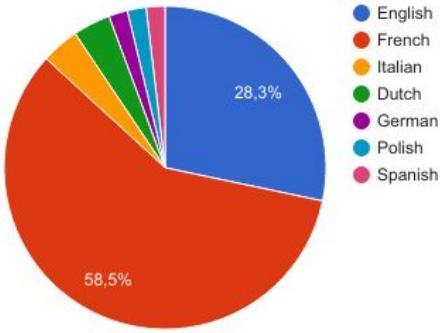


Fig. 9: Distribution of the languages of the users

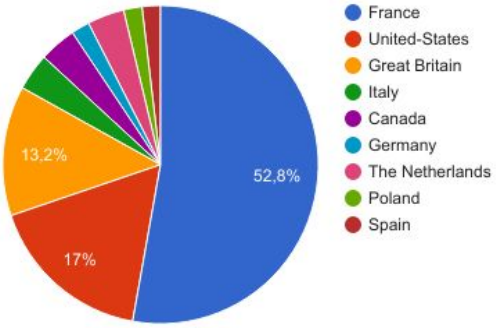


Fig. 10: Distribution of the countries of the users

This distribution of the users reflects the groups targeted on the social media and the efforts made by us on the French community.

5.2.2. Sessions

The users previously presented produced 245 sessions. One session is a set of questions asked to users.
The average of sessions per user is 4.5, but the median of sessions per user is 2. As we said in the section 5.1.3.b. about the bias of engagement, we can see here that some users contributed a lot to the experiment when others did only few sessions.
In this experiment, there are two types of sessions: the “browse” sessions and the “single” sessions. During the experiment, there have been 114 “browse” sessions and 131 “single” sessions.

5.2.3. First experiment: the “browse” sessions

There have been 114 sessions of this type run during the experiment. The first session presented to a user when starting to contribute was of this type. So we can say that this part of the experiment is the one where we collected the most representative data.

Presentation of the questions

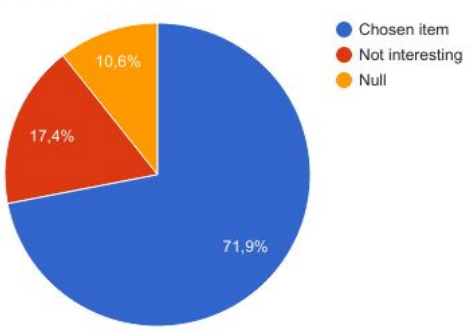


Fig. 10: Distribution of the type of answers for the "Browse" sessions

These 114 sessions generated 309 questions to users. In theory, a session proposed 5 questions to a user. But in many cases, users chose the "No interesting choice" option, which stopped the session (because we couldn't recommend items from a non-choice), or just stopped answering to the questions, which stopped the session.

The answers to the questions are distributed according to the graph "Distribution of the type of answers for the 'Browse' sessions". The "Chosen item" part represents the users choosing an item between the presented items. The "Not Interesting" part is the users considering there is no interesting recommendations for an item. The "Null" are the users which didn't answer to the question.

Distribution of the chosen algorithms in the "browse" sessions

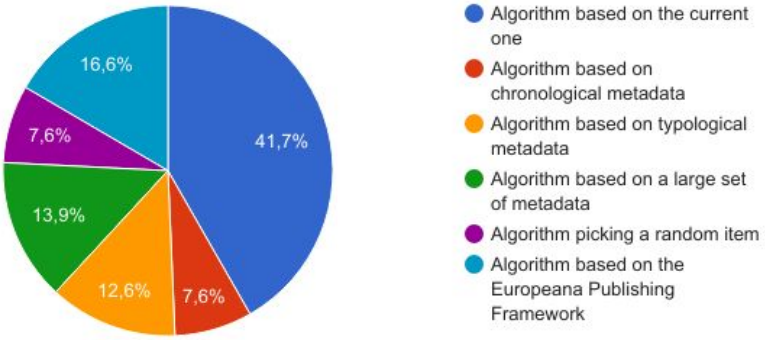


Fig. 11: Distribution of the chosen algorithms in the "Browse" sessions

As described in the section 4.3.1. about the first experiment, each time the users chose an item, they chose one of the algorithm presented in 4.4.3.b.
We present in the following graph the answers to the questions of the "browse" sessions. We can see that the algorithm based on the current algorithm of Europeana Collections is the most popular in the choices of users.
The algorithm based on chronological metadata and the algorithm picking a random item are the less chosen.

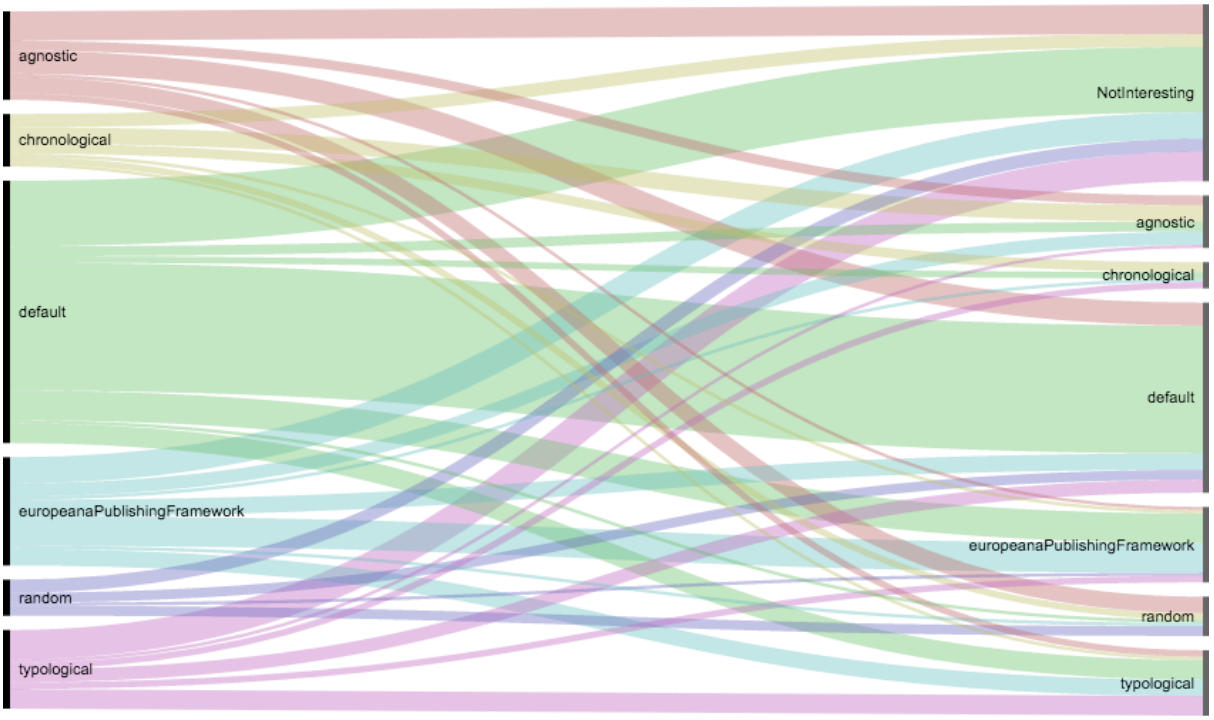
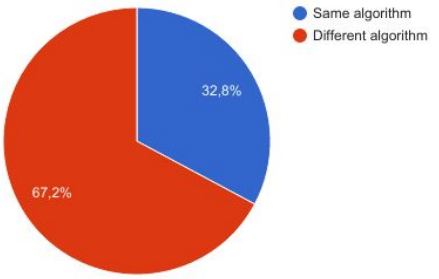


Fig. 12: Distribution of the choices by algorithms according to the previous algorithm

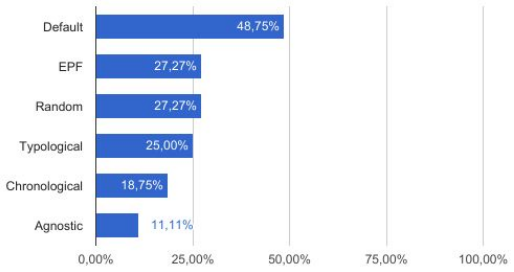
The graph in Fig. 12 represents which algorithm the users chose after browsing a previous item. The left column in the graph represent algorithms for suggested items, which become a reference item after being selected. From this new reference item, we suggest a new set of similar items. In these new similar items, it is interesting to note that the majority of the users chose an item from a different algorithm than for the previous item. A third of the items has been browsed from an item generated with the same algorithm.

Fig. 13: Percentage of clicks from an item to another generated with the same recommender system



The graph in Fig. 14 presents the percentage of chance that the user chooses the item from the same algorithm. We can see here that the algorithm based on the current algorithm of Europeana Collections presents the best score.

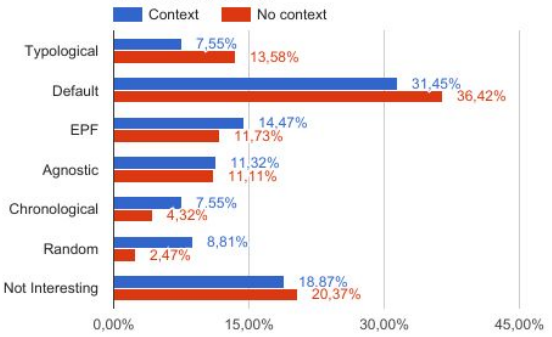
Fig. 14: Ability to continue in the same algorithm



Global distribution of the answers with context criterion in the browse sessions

NB: readers of this section should keep in mind the bias related to contextualization mentioned in section 5.1.3. Plus, in this graph, the column "No Context" profits from a multiplying factor of 1.3641 to compare a similar percentage of sessions with and without context.

Fig. 15: Distribution with and without context of the clicks by algorithm



It presents the variation in the answers when an information of context is given to the user.

It is surprising to realize that the main impact of context is to increase the number of results of the algorithm which picks random items. Other variations seem too small to be taken into account.

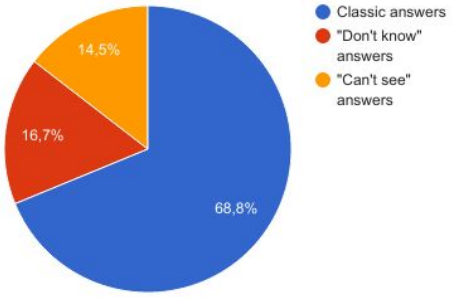
5.2.4. Second experiment: the "single" sessions

There were 131 sessions of this type run during the experiment. These sessions produced 718 questions. 23 questions have a null answer, which means that the user didn't answer and stopped the experiment.

Distribution of the type of answers for the "single" sessions

The answers are distributed in three types: the classic answers, the "Don't know" answers and the "Can't see" answers. The first type (classic answers) represents the user which did rate the recommendation between 1 and 5 as asked by the chatbot. The "Don't know" answers represent the users which clicked on the "Don't know" button. This means they were unable to rate the recommendation, but we don't know exactly why. And the "Can't see" answers represent the users which clicked on the "Can't see" button. This means they were unable to see a part of the recommendation, but we don't know exactly if it was one thumbnail, both, with or without the titles.

Fig. 16 : Distribution of the type of answers for the " single" sessions



Distribution of the answers by algorithm for the single sessions

Fig. 17: Distribution of answers per algorithm

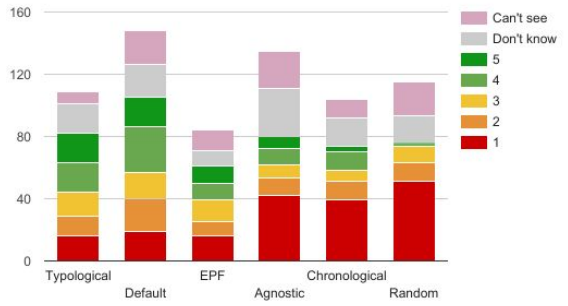
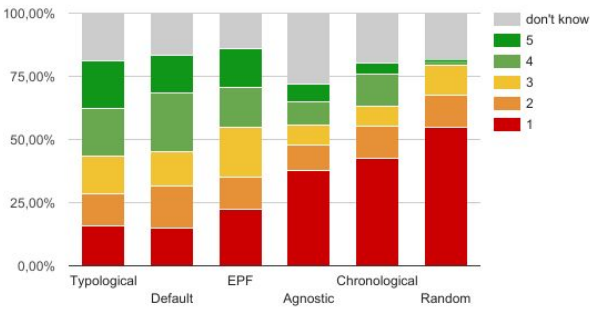


Fig. 18: Distribution in percentage of the answers expressing an opinion per algorithm



The previous graphs present the distribution of the answers by algorithm. The first graph includes the answers "Can't see" as a marker of the data quality / performance of the algorithm of the chatbot. The second graph includes only the answers expressing an opinion. In this last one, we consider the answer "Don't know" as the expression of an opinion. Plus, the second graph expresses the results in percentage. Indeed, the quantity of answers for each algorithm is not related to a choice from users but is a random result caused by the chatbot's computation.

In our analysis, we consider the last graph. Considering the significant but reasonable difference of results, it would not make sense to give more value to algorithms with more answers.

	Typological	Default	EPF	Agnostic	Chronological	Random
1	15,84%	15,08%	22,54%	37,84%	42,39%	54,84%
2	12,87%	16,67%	12,68%	9,91%	13,04%	12,90%
3	14,85%	13,49%	19,72%	8,11%	7,61%	11,83%
4	18,81%	23,02%	15,49%	9,01%	13,04%	1,08%
5	18,81%	15,08%	15,49%	7,21%	4,35%	1,08%
Don't know	18,81%	16,67%	14,08%	27,93%	19,57%	18,28%

Fig. 19: Results in percentage of the answers expressing an opinion per algorithm

We can see that the algorithm based on the typological metadata is the algorithm collecting the biggest number of "5" answers, just before the algorithm based on the current 'similar item' algorithm of Europeana Collections.

If we include the "4" answers as valuable recommendations, the algorithms based on the typological metadata and based on the current algorithm of Europeana Collections are at the same level, and the algorithm based on the Europeana Publishing Framework comes just after.

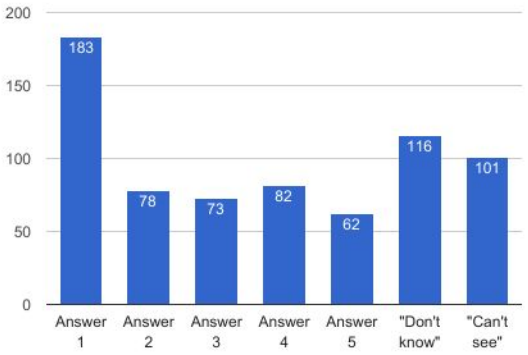
The answers given to the algorithm picking a random item are particularly bad. It demonstrates the value of the recommendation.

Regarding the algorithms based on chronological metadata and on a large set of metadata (the "agnostic" in our graphs), they don't seem valuable for users.

Global distribution of the answers in the "single" sessions

If we don't consider the criterion of the algorithms and we only take into account all the answers, we can notice that the recommendations
are generally not rated as valuable by the users.
Most of the answers are in non-valuable columns: less than 4. We also can consider the "Don't know" and the "Can't see" columns as non-valuable because they don't encourage the user to choose this item.

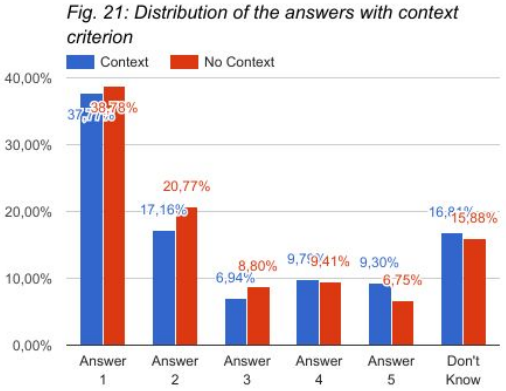
Fig. 20: Distribution of the answers



Global distribution of the answers with context criterion in the single sessions

NB: readers of this section should keep in mind the bias related to contextualization mentioned in section 5.1.3.
It presents the variation in the answers when an information of context is given to the user.
The main result is that there is no significant change in the answer when an information of context is given to the user.
We can see a small inversion of the rating between the negative results (1, 2 and 3) and the positive results (4 and 5) which could give credit to the context information but considering the bias previously exposed and the small quantity of data, it would be a rather bold assumption.

Fig. 21: Distribution of the answers with context criterion



5.2.5: Measures of metadata similarity according to the rates of the users

We have presented in 4.3.4 the following methodology: "*we are going to measure the similarity between metadata fields of items to compute a level or percentage of similarity between items in the recommendations used for the experiment. Then, it should be possible to compute an average of the level/percentage of similarity of the similar items chosen by the users, which I will compare with the general level/percentage of similarity of the items composing the recommendations.*"
We choose to apply a cosine similarity[40] to our data to compute a level of similarity between 0 (no similarity) and 1 (strict similarity, it is the same string). The cosine similarity has been

[40] http://en.wikipedia.org/wiki/Cosine_similarity

applied to the title field (dcTitle in the Europeana metadata), which was the most important and the most visible information during the experiment.

For the results of the "single" sessions, we present in Fig. 22 an average of the cosine similarities per vote. The results highlight that the vote of users increase with the similarity.

Vote	**Average of cosine similarities**	**Number of values**
1	0.05826622776917474	183
2	0.09044263837318006	78
3	0.25683542275958876	73
4	0.40617961588898066	82
5	0.4066716613884886	62
Don't know	0.18405157859045213	116
Can't see	0.12763813088148607	101
Total	**0.18593962184290966**	**695**

Fig. 22: Average of cosine similarities per vote

For the results of the "browse" sessions, we compute for each proposal a sorted list of the suggested items per cosine similarity with the reference item. Then, we compute for each clicked similar item if it was the first, second, third, etc. item in the list sorted by cosine similarity. We present this result in Fig. 23. In this case, the result indicates on the contrary of the previous table, that the cosine similarity would have a decrease effect on the choice of the similar item.

Rank in the sorted list	**Average of the cosine similarity**	**Number of clicked items**
1	0.8386358397227963	18
2	0.6356097150694588	33
3	0.5623688619286644	64
4	0.2603839723573278	31
5	0.16666666666666663	24
6	0.0014471037525422082	53
Total	**0.377626782590578**	**223**

Fig. 23: Number of clicked items per rank of cosine similarity in a proposal list

6. Analysis

6.1. The difficulty of the identification of a logic in a user's selection of recommendations

In the previous results, we can see in the Fig. 13 that only 32.8% of the clicks go from an item generated with an algorithm to another item generated with the same algorithm, in the "browse" sessions. If we consider a simple distribution by chance, it would suggest a percentage of 100/6, meaning about 17%. But in Fig. 14, the high result of the algorithm which picks random item (27.27% of ability to continue the browsing with the same algorithm) makes conclusion hard to draw.
In the Fig. 14 (Ability to continue in the same algorithm), only one algorithm (the algorithm based on the current algorithm of Europeana Collections) is chosen nearly 50% of the time by users after being used for the reference item of the previous stage, which seems to be a really good score.

6.2. The users don't seem to be aware of their logic

The results of the experiment don't show a significate difference of logic from the user with and without information of context given by the chatbot. In some cases, the logic seems to be the opposite of what it should be, for instance when the algorithm picking random items is more chosen when presented context than present without context.
Regarding this section, we should remind the identified bias described in 5.1.3. about the render of Facebook Messenger. It is difficult to say if the given result is due to the bias or if the users didn't give value to the information of context.

6.3. The attractiveness of the current algorithm of Europeana Collections and of its metadata

The most significant conclusion that it seems we can do is the attractiveness of the current algorithm of Europeana Collections, and of its metadata. The algorithm named "Typological" in the results of the experiment uses much metadata of the "Default" algorithm.
These two algorithms are definitively the most chosen between the six algorithms tested. Based on this, we could say that the metadata dcType, dcSubject, dcCreator (shared by the two algorithms) are the favorite of the users.
The Fig. 22 presents the number of objects where these metadata are available[41]. It seems that at the exception of the recommender system based on chronological metadata, the most valuable metadata are also very popular in the objects. For the chronological metadata exception, the reason of the non valuability could be the strict match of metadata required by

[41] Measure done the 04-23-2017 on the test Solr instance of Europeana

the algorithm when an interval (for instance 10 years before and after the date) would be probably better to represent a date.

Metadata	Number of objects	Used in
proxy_dc_title	52 068 269	default, EPF, agnostic
proxy_dc_type	47 436 962	default, typological, EPF, agnostic
proxy_dc_subject	30 263 791	default, typological, EPF, agnostic
proxy_dc_creator	21 427 793	default, typological, EPF, agnostic
proxy_dc_contributor	6 855 776	typological, agnostic
proxy_dcterms_provenance	5 154 146	typological, agnostic
proxy_dcterms_spatial	24 084 859	typological, agnostic
proxy_dcterms_temporal	10 937 202	chronological, agnostic
proxy_dcterms_created	14 553 776	chronological, agnostic
proxy_dc_date	30 688 772	chronological, agnostic
proxy_dcterms_medium	8 471 952	agnostic
proxy_dc_publisher	18 917 515	agnostic
proxy_dc_language	33 049 427	agnostic
proxy_dc_format	18 067 955	agnostic
proxy_dc_description	33 869 134	agnostic

Fig. 24: Number of objects per metadata

This analyze highlights the importance of working on the data quality (as we said, the team of Europeana has engaged this work since few years). But the data quality is not a perfect solution. Indeed, we can't consider data quality as full-completed metadata, as some metadata are not relevant for every object (like proxy_dc_publisher which doesn't concern artworks).

6.4. The users seem to promote the similarity, in the contrast with the serendipity, where there is no point of comparison

We already have presented the different algorithms and highlighted the ability - also its aim - of the current algorithm of Europeana Collections, to find and to recommend the most similar item to the users.
From the fact that this algorithm has been the most chosen in this experiment, it seems that the users promote the strict similarity and not the serendipity that other algorithms (as the

algorithm based on chronological metadata or the algorithm based on the Europeana Publishing Framework) try to accentuate. Our analyze of the cosine similarity in the "single" sessions (Fig. 22) substantiates this hypothesis. On the contrary, the analyze of the cosine similarity in the "browse" session (Fig. 23) highlights that the users prefer non strict similarity. But this result should be taken carefully. Indeed, an analyze of cosine similarity of the recommender system named "default" (Fig. 25) seems to demonstrate that when the users choose the "default" recommender system, they have preferred a medium similarity (in the title). To complete these analyze, it would be necessary to compare these results to the availability of an image for an item, we didn't have this information in our case.

Rank in the sorted list	**Average of the cosine similarity for Default**	**Number of clicked items**	**Total clicked item**
1	1.0	12	18
2	0.6748978121457258	21	33
3	0.5686141832170039	40	64
4	0.34276897528453826	12	31
5	0.6666666666666665	6	24
6	0.0	2	53
Total	**0.6132327858833856**	**93**	**223**

Fig. 25: Number of clicked items per rank of cosine similarity in a proposal list for the recommender system based on the current of Europeana Collections

But this difference between the two experiments could be linked to the presence of points of comparison in the "browse" session. It would be necessary to do this experiment again but with suggested items from the same recommender system to evaluate this assumption.
Finally, it is important to highlight that the data quality may influence this result. Indeed, it is difficult for a user to make choices when a large amount of data doesn't have thumbnail or title. The users probably looked for safety in their choices because they considered that they didn't have enough information to choose.

6.5. Facebook Messenger and chatbot, the power of a large group of users and of the natural language, but with new biases

Two of the four biases we addressed in the section 5.1.3. are due to Facebook Messenger. The bias of the context and the bias of the rendering of the thumbnail have hindered the scientific value of our evaluation. Indeed, it is difficult to render a strict architecture of the information on Facebook Messenger because it is no possible handling of the style of the text.

In our case, the ability to render the text of the context in italic, for instance, may resolve our bias.
But Facebook provides us a large group of users, and moreover, users with a higher level of contribution than we may get via a classic survey. Indeed, the chatbot allows reminders and many users have come back several times.
Plus, it appears that people have interacted with the bot, trying to detect whether it performs some Natural Language Processing (NLP) (even if there is not NLP in this bot). People tried to "play" with the bot, trying to see if the bot was going to change its answers according what they said (like Siri, Cortana, etc.). Interactions between the users and the system was higher than with a classic survey, and the users seemed sensitive to the natural language of the bot.

7. Conclusion

The work done during the time of our thesis should be continued to go deeper in the question of the preferences of the users about the recommendations of cultural heritage objects in the digital library Europeana. First, it would be useful to test a recommender system with data with semantic links. Second, it would be necessary to include the new logging framework of Europeana in the recommender system and to assess it.
Indeed, our work highlights the difficulty of making recommendations without user profile and the importance of the data quality in the field of the recommender systems.
The improvement of the recommender system in Europeana Collections is a large work which is linked to the general question of the information retrieval in Europeana. In our experiment, the results of some specifics recommender systems (as the recommender system based on the Europeana Publishing Framework) point the importance of the development of a recommender system which will take into account every CHO and not only the most beautiful or reusable. Indeed, the great heterogeneity of the collection of Europeana makes difficult to limit the recommendation to only a part of it, on the contrary of what we thought at the beginning of the work.
We definitely think that the improvement of this recommender system will come from the points we previously highlight and considering the work done by the team of the Europeana Foundation, we believe the recommender system will increase soon.

8. Glossary

- **CHO**: acronym of 'Cultural Heritage Object', this is an item in Europeana Collections.
- **More Like This**: name of the recommender function in the source code of Europeana and in Solr.
- **Recommender system**: the algorithm providing related items. This is a subclass of information filtering systems.
- **Record**: The package of data about a CHO (i.e. the CHO enriched by different web resources and some provenance data)
- **Reference item**: we consider the reference item as the item from which related items are computed.
- **Related item**: refers to every distinct object (physical or not) from which we can establish a relation with a CHO by similar properties, user ranking or other.
- **Similar items**: heading of the recommended items on Europeana Collections' current record page.
- **Content**: In the Europeana context, content is understood as a digital representation of a cultural heritage object. So content is most of the time the web resource of a CHO, using the notions of the Europeana Data Model. But in this thesis, we use the term "content" to define any group of resource, data or metadata constituting an information, e.g. an entity. We justify this choice by the common acceptation of the expression "content-based recommender system", which refers to our definition. When we will need to talk about Europeana content, we will use the expression "web resource".
- **End-users:** We understand "end-users" here as users who "typically do not possess the technical understanding or skill of the product designers" such as in [https://en.wikipedia.org/wiki/End_user, 04/14/2017]. Here, they are the customers of the products based on Europeana Collections.

More is available in [Europeana Data Model - EDM Mapping Guidelines, 2016].

9. Bibliography

Aletras, Stevenson, and Clough, 'Computing Similarity Between Items in a Digital Library of Cultural Heritage', *J. Comput. Cult. Herit.*, 5 (2013), 16:1–16:19 <https://doi.org/10.1145/2399180.2399184>

Charle, 'Europeana Data Model Documentation', *Europeana Professional*, 2014 <http://pro.europeana.eu/page/edm-documentation> [accessed 27 December 2016]

Clough, Otegi, Agirre, and Hall, 'Implementing Recommendations in the PATHS System', in *Theory and Practice of Digital Libraries -- TPDL 2013 Selected Workshops*, ed. by Łukasz Bolikowski and others, Communications in Computer and Information Science (presented at the International Conference on Theory and Practice of Digital Libraries, Springer International Publishing, 2013), pp. 169–73 <http://link.springer.com/chapter/10.1007/978-3-319-08425-1_17> [accessed 16 December 2016]

'Definition of the Europeana Data Model v5.2. 7' <http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.7_042016.pdf> [accessed 28 December 2016]

'Europeana Data Model – Mapping Guidelines v2. 3' <http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.3_112016.pdf > [accessed 27 December 2016]

Europeana foundation, 'Europeana Strategy 2020: "We Transform the World with Culture"', 2015 <http://strategy2020.europeana.eu/> [accessed 28 December 2016]

Hill, Charles, and Isaac, 'Discovery and Metadata - Scenarios - v2', *Google Docs* <https://docs.google.com/document/d/1ej0ouDg_uhOVnE1LE2-IEtI9xNhMpeqzNIIwSjLoAbI/edit?usp=embed_facebook> [accessed 16 December 2016]

de Hoog, 'How to Share Data', *Europeana Professional*, 2014 <http://pro.europeana.eu/page/how-to-contribute-data> [accessed 23 April 2017]

Iaquinta, Gemmis, Lops, Semeraro, Filannino, and Molino, 'Introducing Serendipity in a Content-Based Recommender System', in *2008 Eighth International Conference on Hybrid Intelligent Systems*, 2008, pp. 168–73 <https://doi.org/10.1109/HIS.2008.25>

Lops, Gemmis, and Semeraro, 'Content-Based Recommender Systems: State of the Art and Trends', in *Recommender Systems Handbook*, ed. by Francesco Ricci and others (Springer US, 2011), pp. 73–105 <http://link.springer.com/chapter/10.1007/978-0-387-85820-3_3> [accessed 28

December 2016]

Murphy, '2016: Top 20 Searches on Europeana | Europeana Blog', *Europeana Blog* <http://blog.europeana.eu/2016/12/2016-top-20-searches-on-europeana/> [accessed 27 December 2016]

Oufaida, and Nouali, 'Exploiting Semantic Web Technologies for Recommender Systems: A Multi View Recommendation Engine', in *Proceedings of the 7th International Conference on Intelligent Techniques for Web Personalization & Recommender Systems - Volume 528*, ITWP'09 (Aachen, Germany, Germany: CEUR-WS.org, 2009), pp. 87–92 <http://dl.acm.org/citation.cfm?id=2907132.2907142> [accessed 16 December 2016]

Pekel, 'How to Share Data - Europeana Professional' <http://pro.europeana.eu/share-your-data/how-to-contribute-data> [accessed 27 December 2016]

Pineau, 'BabelNet Assessment (French Edition) for Europeana Enrichment', *Google Docs*, 2016 <https://docs.google.com/document/d/1tgaN_k0U2XjDSxVxEFDa-bq3d5GmnpXsd-a5xxXfy5g/edit?usp=embed_facebook> [accessed 24 December 2016]

———, 'Dataset for Evaluation of Europeana Data', *Google Docs*, 2016 <https://docs.google.com/document/d/11lJqOlGbTkGwPOokUMotp8zPzEnicSXhMboHaoAHKFs/edit?usp=embed_facebook> [accessed 24 December 2016]

———, 'Evaluation of the Current Europeana "More Like This" Function', *Google Docs*, 2016 <https://docs.google.com/document/d/1bBVlbRMk21yUyFfX-Ate8Dg80YKEl3gUyYFJi82JxKM/edit#heading=h.3ruk5rmwshqw> [accessed 24 December 2016]

———, 'Experimental Protocol for Query of Similar Items', *Google Docs*, 2016 <https://docs.google.com/document/d/1UZ-WhmqnsDmzbdDygh0RBjhcEBRpytDVyzLyRSfryvg/edit?usp=embed_facebook> [accessed 24 December 2016]

———, 'Multilinguality in the MLT Function of Europeana', *Google Docs*, 2016

<https://docs.google.com/document/d/10EBVAVpk6kCGh4tZ60aIZxDl5eS50ivObEJYwv1QXBA/edit?usp=sharing> [accessed 19 December 2016]

———, 'Score of the Results in the Similar Items Query of Europeana Recommender System', *Google Docs*, 2016 <https://docs.google.com/document/d/1SvL1Px_xzzU9QdhZRMsR2YWPEjABMZ-FREQL3x1T7fM/edit> [accessed 24 December 2016]

Resmini, and Rosati, *Pervasive Information Architecture - 1st Edition* <https://www.elsevier.com/books/pervasive-information-architecture/resmini/978-0-12-382094-5> [accessed 18 February 2017]

Schafer, 'DynamicLens: A Dynamic User-Interface for a Meta-Recommendation System', in *Beyond Personalization 2005* (presented at the Inte rnational C onf erence on Intelligent User I nterfaces, San Diego, 2005) <http://grouplens.org/beyond2005/bp2005.pdf#page=72>

Schafer, Konstan, and Riedl, 'Meta-Recommendation Systems: User-Controlled Integration of Diverse Recommendations', in *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, CIKM '02 (New York, NY, USA: ACM, 2002), pp. 43–51 <https://doi.org/10.1145/584792.584803>

Shen, 'Design for User Engagement on Europeana Channels' (Delf Univerity of Technology, 2014) <https://drive.google.com/file/d/0B5Cq8nCz41blck5qTldtSUwtUDg/edit?usp=embed_facebook> [accessed 18 April 2017]

Son, 'Dealing with the New User Cold-Start Problem in Recommender Systems: A Comparative Review', *Information Systems*, 58 (2016), 87–104 <https://doi.org/10.1016/j.is.2014.10.001>

Wang, Aroyo, Stash, and Rutledge, 'Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case Study', in *User Modeling 2007*, ed. by Cristina Conati, Kathleen McCoy, and Georgios Paliouras, Lecture Notes in Computer Science (presented at the International Conference on User Modeling, Springer Berlin Heidelberg, 2007), pp. 385–89

<http://link.springer.com/chapter/10.1007/978-3-540-73078-1_50> [accessed 16 December 2016]

Wang, Wang, Stash, Aroyo, and Schreiber, 'Enhancing Content-Based Recommendation with the Task Model of Classification', in *Knowledge Engineering and Management by the Masses*, ed. by Philipp Cimiano and H. Sofia Pinto, Lecture Notes in Computer Science (presented at the International Conference on Knowledge Engineering and Knowledge Management, Springer Berlin Heidelberg, 2010), pp. 431–40 <http://link.springer.com/chapter/10.1007/978-3-642-16438-5_33> [accessed 16 December 2016]

10. Appendices

Appendix 1: The chatbot developed for the experiment

The code sources (all free to reuse) of the chatbot are available at the following addresses:

- The Python script developed for the computation of the recommendations before starting the experiment: Github repository
- The API developed under the framework Symfony3 (PHP) for the feeding of the chatbot in data: Github repository
- Plus, the scenario of the chatbot has been developed with the web-service Chatfuel

The chatbot is available here:

- On Facebook Messenger: by searching "Europeana Evaluation chatbot"
- On Facebook: on the "Europeana Evaluation chatbot" page

Playing with the chatbot requires a Facebook account.

The administration page of the API with the data collected are available on a dedicated website.

Appendix 2: Results of the experiment

The results of the experiment can download at the following address.

Appendix 3: Queries of the tested recommender systems in the experiment

- **Name:** Recommender system derived from the current recommender system
- **Short Name:** default
- **Query rules:** Python function on Github
- **Example:** for this reference item , the related items query is:

```
(what:("szobor")
OR who:("http://viaf.org/viaf/24604287")
OR proxy_dc_title:("Lovas")
OR provider_aggregation_edm_dataProvider:"Szépművészeti Múzeum")
AND NOT europeana_id:"/2032004/2778
```

- **Name:** Recommender system based on a large set of metadata
- **Short Name:** agnostic
- **Query rules:** Python function on Github
- **Example:** for this reference item, the related items query would be:

```
(("Lovas")
OR ("bronz")
OR ("2032004_Ag_EU_EInside_SzepmuveszetiMuzeum-FAB")
OR ("hu")
OR ("Szépművészeti Múzeum")
OR ("EUInsideDA")
OR who:("http://viaf.org/viaf/24604287")
OR what:("szobor")
OR when:("1501-1550"))
AND NOT europeana_id:"/2032004/2778"
```

- **Name:** Recommender system with selection of metadata
- **Short Name:** typological
- **Query rules:** Python function on Github
- **Example:** for this reference item , the related items query is:

```
(what:("szobor")
OR who:("http://viaf.org/viaf/24604287"))
AND NOT europeana_id:"/2032004/2778"
```

- **Name:** Recommender system with chronological metadata
- **Short Name:** chronological
- **Query rules:** Python function on Github
- **Example:** for this reference item , the related items query is:

```
(when:("1501-1550"))
AND NOT europeana_id:"/2032004/2778”
```

- **Name:** Recommender system based on the Europeana Publishing Framework
- **Short Name:** europeanaPublishingFramework or EPF
- **Query rules:** Python function on Github
- **Example:** for this reference item , the related items query is:

```
(what:("szobor")
OR who:("http://viaf.org/viaf/24604287")
OR title:("Lovas")
OR DATA_PROVIDER:"Szépművészeti Múzeum")
AND NOT europeana_id:"/2032004/2778"
&qf=IMAGE_SIZE:medium&qf=IMAGE_SIZE:large&qf=IMAGE_SIZE:extra_large
&thumbnail=true
```

- **Name:** Recommender system with random item
- **Short Name:** random
- **Query rules:** Python function on Github