POLITECNICO
MILANO 1863

DIPARTIMENTO DI DESIGN

Displaying open cultural collections. Design guidelines for cultural content aggregators.

PhD candidate
Giovanni Profeta

Supervisor
Professor Paolo Ciuccarelli

Co-supervisor
Michele Mauri

External examiners
Professor Isabel Meirelles, Professor Silvia Miksch

Head of the Doctoral Program
Professor Paola Bertola

“La collezione nasce del bisogno di trasformare lo scorrere della propria esistenza in una serie di oggetti salvati dalla dispersione, o in una serie di righe scritte, cristallizzate, fuori dal flusso continuo dei pensieri”.

“The collection was born out of the need to transform the flow of your existence into a series of objects saved from dispersion, or into a series of written, crystallized lines, out of the continuous flow of thoughts”

Italo Calvino

Acknowledgements

I would like to thank my supervisors Prof. Paolo Ciuccarelli and Michele Mauri for their advice and support. I also would like to acknowledge my colleague Iolanda Pensa for her feedback and suggestions.

I thank the coordinator of the PhD in Design of Politecnico di Milano, Prof. Paola Bertola, the PhD committee, in particular Prof. Luca Guerrini and Prof. Raffaella Trocchianesi, and the secretary of the PhD in Design Simona Murina.

A Sspecial thanks to my colleagues of DensityDesign research lab for the ideas and knowledge they share with me: Maria De Los Angeles Briones Rojas, Gabriele Colombo, Matteo Azzi, Andrea Benedetti, Alessandra Facchin, Eleonora Cappuccio, Daniele Cimineri, Tommaso Elli, Lea Fabiano, Beatrice Gobbo, Michele Invernizzi, Marco Pappalepore, Azzurra Pini, Ginevra Terenghi and Giorgio Uboldi.

Thanks also to my PhD colleagues: Silvia Franceschini, Alice Devecchi, Elisa Bacchetti, Mariana Fonseca Braga, Marta Corubolo, Maria Renata Guarneri, Agnese Piselli, Marta Isabella Reina, Maria Elena Soriero, Bogdan Stojanovic, Isa Helena Tiburcio, Ruta Valusyte and Anna Serena Vitale.

I thank my colleagues of the Laboratory of visual culture of the University of Applied Sciences and Arts of Southern Switzerland (SUPSI): Massimo Botta, Jean-Pierre Candeloro, Serena Cangiano, Silvia Converso, Vanessa De Luca, Florence Devouard, Davide Fornari, Fabian Frei, Elisabetta Lazzaroni, Marco Lurati and Marta Pucciarelli.

I also would also like to express my appreciation to Marian Dörk, head of the Urban Complexity Lab at University of Applied Sciences Potsdam, and all the team members, for their advice and for having hosted me during the development of the design project.

With regard to the design projects, I would like to thank the Wikimedia Foundation, Wikimedia Switzerland, Wikimedia Italia, Wikimedia Germany, and my Wikipedian friends. I also thank Juliane Stiller for her feedback and suggestions about the methodology and the results of my thesis.

Last, but not least, I thank my family, in particular my sister Antonella Anna, and my friends from Lugano for the support.

Publications

During the PhD studies, the following papers and articles have been published.

Menichinelli, M., Profeta, G., Cangiano, S. (2020). "Visualizing open communities. Guidelines from three case studies", in *Proceedings of 2CO COmmunicating COmplexity conference 2018*, University of La Laguna, Tenerife.

Cannata, M., Profeta, G., Voegeli, M., Lüscher, M., Morandi, L. (2018). "Crossing SSH and STEM approaches in a MapDesign course using open data and software". *PeerJ Preprints* 6:e27237v1 https://doi.org/10.7287/peerj.preprints.27237v1.

Profeta, G. (2018) "Metodi visuali per esplorare l'impatto delle istituzioni culturali su Wikipedia", in *FRID Fare Ricerca In Design*, pp. 453-463, Mimesis, Milan-Udine.

Botta, M., Profeta, G. (2018) "Towards the Information Visualization of Connected Objects" in Luca Guerrini (ed.), Dialogues on Design. Notes on Doctoral Research,pp. 113-122, Franco Angeli, Milan.

Profeta, G. (2017, May) "The graphic visualisation of flows. The information design tools for the analysis of the urban fabric". *Flows. Modelling Mobility*.

Cangiano, S., Profeta, G., Lurati, M., Elsharbaty, S., Pensa, I.& Devouard, F. (2017, February) Community Digest: Using data to visualize Wikipedia knowledge gaps", Wikimedia Foundation Blog.

Profeta, G. (2016, December) Osservare la città connessa. Strumenti visuali per la mappatura di dinamiche quotidiane, in *MD Journal* (2), pp. 134-145. ISSN 2531-9477.

Content

List of figures

List of tables

Note
The beginning of every chapter displays images of projects based on digitised cultural content and collections that are released in the public domain or open license. Most of the featured projects contributed to the advancement of humanistic, artistic and technological research. The selection of these projects intends to highlight the relevance of digitalisation and free release of cultural content as a means to foster the production of new knowledge.

Abstract

In recent years, archives, museums and other cultural institutions have introduced several strategies to make their collections more accessible. The digitisation of extensive cultural collections and their release under open licenses are fostering the creation of cultural content aggregators, web platforms gathering several collections which support scholars, writers and artists in their research, dissemination and artistic activities. In comparison with digital archives, which show a single, curated collection, cultural content aggregators have a more complex information architecture consisting of several collections.
Although cultural aggregators are adopting more performative technologies and rigorous sharing methods, their user interfaces have several usability issues. Thus, most of the available heritage is invisible to the end-user, like it is in a sort of a digital depot.
This work investigates classification systems and interface solutions that may foster the access and the use of digitised cultural objects on cultural content aggregators.
The research has been conducted according to a research through design approach. The best practices for designing classification systems and interface solutions are based on an in-depth analysis of existing aggregators and the development of an interface prototype, which is validated by the primary stakeholders.
The results show that facilitating the access and the use of cultural objects does not require only technological advancement, but also a shift in the way the cultural content aggregator is intended. It no longer has the mere function of a searchable database, but it is acquiring the role of a research and dissemination assistant.
The thesis presents the results as design guidelines based on the OpenGLAM principles listed by the Open Knowledge Foundation and the European Community - a set of guidelines that aim to support the release of open contents. The design guidelines aim to provide designers with a tool which could support them in the whole design process, including the design of the information architecture and interface elements for using and sharing content.

Introduction

Museums, libraries, archives and other cultural institutions have been digitising and providing access to their collections for years. This digital transformation is due to several reasons regarding, in particular, the physical limit of the exhibit spaces and the request for remote access to digitised collections by users. In this context, several information systems have been developed in order to allow users to search and access the catalogue of cultural collections.
Recently, cultural institutions have released part of the digitised cultural objects under open licenses in order to increase the usage and the awareness of their cultural collections.
The availability of open cultural collections, together with richer metadata, is fostering the design and implementation of web-based applications which provide digital access to resources belonging to collections that are physically disconnected. These cultural content aggregators are particularly conceived for students, scholars and writers (in particular, journalists and bloggers) who need digitised artworks for educational, research, artistic or dissemination purposes.
From the literature review and an analysis of the existing cultural content aggregators gathering European digitised collections, it emerges that most of them have several limits in terms of content and usability. These issues may reduce the possibility to access digitised collections. Thus, the thesis was born from the observation of the contrast between the will of cultural institutions to open part of their collections and the access limits concerning cultural content aggregators.
This work investigates classification systems and interface solutions that may foster the access and the use of digitised cultural objects on cultural content aggregators. It concentrates on aggregators which collect digital surrogates of cultural objects, with a specific focus on digital images and their metadata. The aim of this thesis is to contribute to the understanding of the dynamics generated by digitised cultural objects openly released on content aggregators and to the design of more effective user interfaces. My contribution is at the intersection between design and Digital Humanities. In particular, it tries to integrate the design research on the access to digitised collections with the emerging interaction modalities which are taking place thanks to open collections. This work does not only present interface features for the fruition of digitised collection, but also introduces classification methods.

Research questions

The research investigates the use of interface tools to foster the access and use of digitised collection on cultural content aggregators. It is framed within the European context, with its regulations and content aggregators. The research relies on the idea that visual tools can be used both as an access point to the collection and an analytical tool to enhance the quality of the content. The research questions are the following:

- Which interface characteristics can foster the access and the use of open digital collections?
  - Which information architectures can enhance the findability of digital surrogates?
  - Which design strategies can facilitate the use of digital surrogates?
- Can we apply the design features used by digital archives to cultural content aggregators interfaces?

In order to answer the first question, I analysed four aspects which characterise conventional web-based information systems: classification, access, navigation and usage of the content. Classification refers to the modalities administrators and end-users adopt to organise the content, access concerns how end-users reach the content, navigation shows the modalities to move among contents and usage refers to the way end-users interact with the content (creation, editing and organisation).
I examined the four aspects through stakeholders and end-user interviews, together with the analysis of existing cultural content aggregators. Furthermore, through the development of the Map the GLAM project, I investigated a case study of how an open collection takes shape and spread within a content aggregator.
In order to answer to the second question, I analysed the existing literature about the interaction with cultural digital archives. With the information gathered, together with the results coming from the methods I used to reply to the first question, I tried to validate the usefulness of a set of features related to the access and use of digitised collections. Thus, I designed and prototyped GLAM Culture Hub, a cultural content aggregator that has been tested by end-users.

Thesis structure

The thesis consists of six chapters. The following list briefly presents the content of every chapter.

First chapter: context

This chapter presents the context in which the thesis is framed, taking into consideration a series of historical facts which brought to the introduction of open licenses in the cultural sector. It also introduces stakeholders and end-users of cultural content aggregators and the terminology adopted in this field.

Second chapter: methodology

This chapter presents the methodological path I used to reply to the research questions, focusing on the three main aspects I developed (the scientific research, the empirical research and the outcome) with their related methods. It also introduces research projects and their goals.

Third chapter: state of the art

This chapter shows the current approaches in designing user interfaces for cultural collections. Through the examination of the specific literature and case studies, it examines the current state of the art regarding the design of cultural content aggregators. The chapter also provides an overview of the methods used by different typologies of cultural institutions to classify and organise their collections.

Fourth chapter: design projects

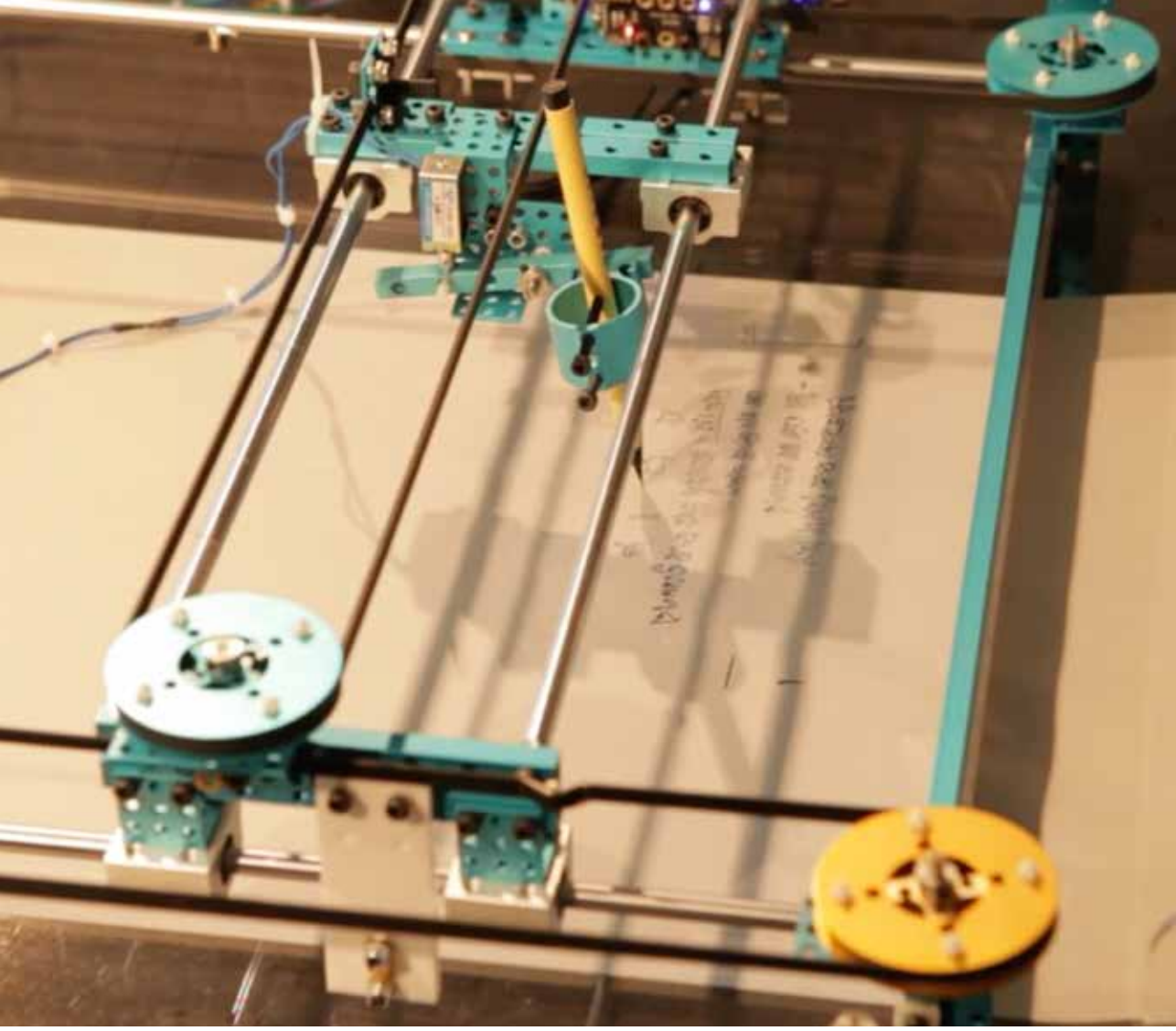
This chapter focuses on the design of two research projects: an analysis of real digital surrogates within an existing cultural content aggregator and the design of a new one. It presents the design processes, the findings and the discussion of the results.

Fifth chapter: design guidelines

This chapter provides a set of five design guidelines concerning how to foster access, navigation and usage of digitised collections within cultural content aggregators. The guidelines are conceived as a complementary tool for the existing OpenGLAM principles, a set of guidelines for cultural institutions developed with the support of the European Commission.

Sixth chapter: conclusion

In this chapter, I recap my contribution to the design research on digitised cultural collections by discussing the overall results I gained through the analysis and presenting the strengths and weaknesses of the research. It also suggests further studies.



The Marginalia Machine
by Tom Schofield, 2014
Website: http://cargocollective.com/tomschofieldart/The-Marginalia-Machine

The Marginalia Machine is a drawing machine that reproduces editorial notes from the Bloodaxe Archive, an archive of contemporary poetry located at the Newcastle University. The artist developed a computer vision software to separate the notes from the background text as new documents are digitised. The extracted notes are then drawn publically to a continuous paper scroll through a Cartesian plotter. This artwork is part of the project "The Poetics of the Archive", a research project that aims to create exploratory interactions with the Bloodaxe Books archive, both online and in physical space. The content of the archive can be used for non-commercial purposes only.

1

Open collections and cultural content aggregators

Nowadays, one of the primary responsibilities for a cultural institution, such as a museum or a library, is to promote scientific research and education. Several cultural institutions are digitising part of their collections and releasing the related digital outcomes under open licenses in order to fulfill this task.

Among the cultural institutions which are realising open cultural objects, there are the Rijksmuseum in Amsterdam and the Museum of Modern Art in New York (MoMA). The Rijksmuseum has released over 600.000 images through Rijksstudio (see tab. 1.1), a web application that makes part of the museum's collection available free of charge. It means that everyone is allowed to copy, modify and distribute the artwork, even for commercial purposes, without asking permission. The MoMA realised over 370.000 images in Public Domain and data on over 125.000 works into the Public Domain by posting them on GitHub[1]. The museum is also collaborating with Google Arts and Culture with the hope to encourage users to generate new contents inspired by the collection (Paqua, 2018).

This kind of experiences have shown several opportunities related to the adoption of open licenses. Among them, there is the increase invisibility for the cultural institution, both as a resource for a specific cultural sector and as a distinctive brand identity, as well as a growth of possible collaborations with other partners using open licenses.

National and international reports widely document the phenomenon of the opening of cultural collections. Accord-

ing to the "PSI in the Cultural Sector[2]", a directive provided by the European Commission, 75% of the significant cultural institutions offers free access to their digitised contents to end-users, while the remaining part charges for licences. According to ENUMERATE survey[3], promoted by Europeana and co-financed by the European Union, in 2017, 58% of the digital objects managed by the institutions was available online. The most "open" institutions in terms of releasing cultural items and related metadata were the libraries, with 58% of collections and 76% of metadata, respectively, available online for general use. According to the survey, the institutional website, usually meant as a digital archive, represents the main channel to access the collection (see fig. 1.1). Other relevant channels are social media platforms, API services and cultural content aggregators. The development of cultural content aggregators was favoured by the introduction of open licenses.

Even though digital archives and cultural content aggregators manage the same kind of content, they have several differences which is important to mention for the comprehension of the specificity of this work.

A cultural digital archive is a repository that stores one or more collections, owned by a cultural institution, to provide long-term access. The main end-users of digital archives are scholars and casual users. They usually search for specific

Tab. 1.1
Tab. 1.1: Collections released under open licenses by cultural institutions. Data gathered from the institutional websites in October 2019.

cultural institution	typology of content	quantity	license
British Library	High-resolution images of cultural objects	Around 1.000.000	Public Domain (CC0)
Museum of Modern Art	Museum of Modern Art High-resolution images of cultural objects	Around 375.000	Public Domain (CC0)
Rijksmuseum	High-resolution images of cultural objects	Around 600.000	Public Domain (CC0)
The British Museum	High-resolution images of cultural objects	Around 1.000.000	CC-BY-SA
V&A Collections	High-resolution images of cultural objects	Around 800.000	some usage restrictions (non-commercial)

contents or enjoy the collection by browsing it. End-users generally benefit from the digitised collection in terms of online visualisation and publication on digital and paper support (when it is allowed).
A curator manages the digital archive by selecting and organising the items in a virtual space. This time-consuming activity has the goal to make the digitised collection accessible and to allow scholars to find novel insights.
Within digital archives, collections take shape not only from a curator but also from several technical reasons, as well as administrative and institutional issues (Krautly, 2016). The reuse of content can be free, bound or limited.

A cultural content aggregator is a repository that stores multiple collections, thanks to the contribution of cultural institutions and the user community, to facilitate the discoverability of the collections. It aims to provide inspiration", in the sense of "creative stimulus", to end-users (Hill et al., 2016). Cultural content aggregators were born from the initiatives of nonprofit organisations, after the introduction of digital archives. A content aggregator aims to promote crowdsourcing, education and entertainment across multiple collections.
The main end-users of content aggregator are scholars and communities of people. Their motivation is based on their interest on the topic and on their will to give a meaning-

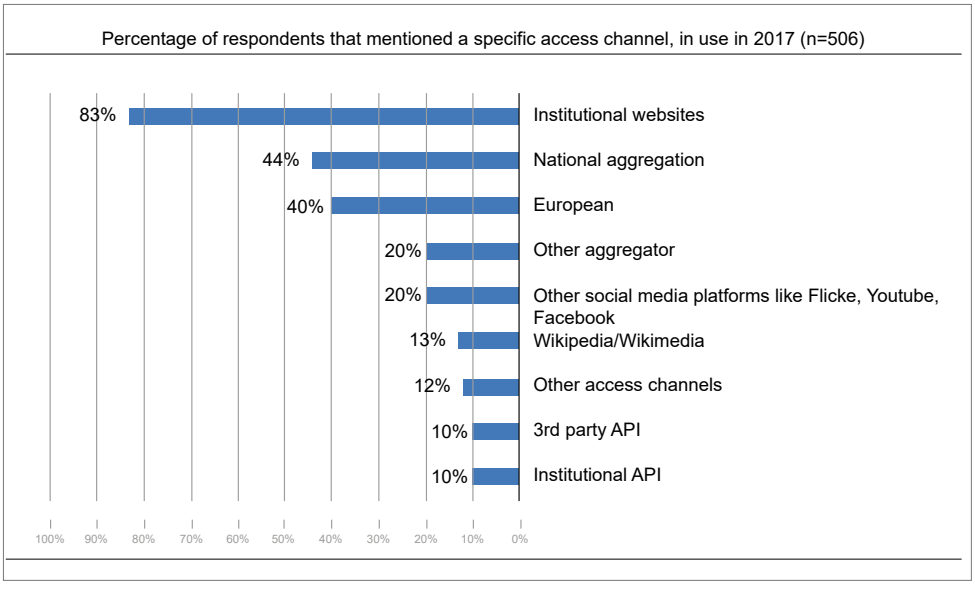


Fig. 1.1
Access channels, according to ENUMERATE report.

ful contribution to the open digital platform. End-users not only visualise and share digitised objects but also use the digital copies to make derivative artworks.
Content aggregators use aggregation logic to gather and display contents provided by cultural institutions and by the community. The reuse of content can be free (mostly) or limited.

1.1 From preserving to opening collections

Since galleries, libraries, archives and museums (GLAMs, for short) were born, their main task was to collect, interpret and preserve their cultural collections.
The first libraries and archives were introduced around 2600 BC. They collected clay tablets in cuneiform script, the earliest form of writing (Casson, 2002). Most of the collections remained private for centuries. Only a few libraries, such as the Library of Alexandria (the most extensive library in the ancient world), in Egypt, allowed scholars or an educated public, to access the papyrus scrolls (Phillips, 2010).
In the 8th century, with the introduction of the papermaking techniques, bibliographic works started to be accessible both for consultation and loan to élites of people (generally, they were members of a particular school or library).
In the 16th century, following the invention of the printing with movable type by Johannes Gutenberg, books rapidly spread in Europe and worldwide. Thus, they become more and more accessible to the general public.

The first known collection of objects was opened at the University of Alexandria in the 3rd century B.C. (Phillips, 2010). The history of collections of objects up to the Middle Ages presents many gaps due to the scarcity of documentation. Around the 14th century, in the northern European countries “Wunderkammer”, or “wonder-rooms”, emerged. They were rooms containing objects belonging to natural history, ethnography, archaeology and religious relics (Fiorio, 2011, p. 10).
Between the 14th and 15th century the idea of the “studiolo” was introduced in Italy. The “studiolo” is a place conceived not only for intellectual activity but for the conservation of artworks, books and other objects with a cultural value. These objects could belong to the past or be commissioned by the owner. Generally, “studioli” were private, extensive collections of objects. They were born from élites of intel-

lectuals, interested in collecting items, that aim to gain prestige, power and wealth. Within these spaces, the arrangement and sequence of objects were closely related to the surrounding environment (Fiorio, 2011, p. 12).
In the 16th century, even dynasties of nobles dedicated large spaces to their prestigious collections of objects. They commissioned architects to contribute to a better presentation of the objects. Thanks to the nobles it was introduced the idea that the exhibition of artworks is as important as the acquisition (Fiorio, 2011, p. 18). For this reason, and for the physical limitations of the exhibition space, only a part of the cultural objects are exhibited (the remaining part is kept in deposits) with the aim of guaranteeing the quality of the visit (Marini Clarelli, 2005, p. 13).
Noble families built or redesigned buildings to exhibit their collections of artworks. A noteworthy building converted into a museum are the Uffizi, in Florence. They were designed by Giorgio Vasari in 1560 as administrative offices but few years later Francesco I de ʻMedici converted them into exhibition rooms (see fig. 1.2). In the same period, the term "musaeum[4]" was used as the equivalent of "collection" (Fiorio, 2011, p. 18).
In the 17th century, during the Enlightenment period, the museum was recognized as a national institution due to the nationalization of royal collections and the confiscation of aristocratic and ecclesiastical property. The museum is recognized as an institution of national interest. Thus, the state took charge of its administration, and access to the collections became a citizen's right (Florio, 2011).
In the 19th century, the museum became a structure able to enhance the cultural heritage of the territory and to increase its economy (Branchesi, 2016). It has also become relevant for scientific research, educational purposes and community outreach (Roberts, 1997).

In the 1960s, following the advent of the computer and digitisation technologies, cultural collections coming both from libraries and museums started to be managed and accessible via computer-based information systems. This happened because of the need to easily manage the growing collection. The traditional paper-based practices (including cross-referencing items and maintaining up-to-date records) became time-consuming and overwhelmed by the collection growth. Furthermore, users started to demand remote access of cultural objects. The majority part of objects

without their counterpart in an information system was not used (Williams, 2010).
In the 1990s, when the Internet reached the mainstream, digitised cultural collections have become accessible without restrictions of time and space. Most institutions had the chance to revive the archived heritage by publishing its digital version on their institutional website or on a web-based digital archive. The Web allowed scholars to develop new ways to analyse the cultural object. Furthermore, thanks to the increase of the demand for more efficient information systems and the need for a more accurate record-keeping, new professionals were born.

In the 2000s, following the introduction of free licenses of use (in particular the Creative Commons licenses), cultural institutions are freely licensing part of their digitised collections. Several projects have been introduced with the aim to support cultural institutions in the digitisation and the opening of their collections. Among these, there are OpenGLAM, Cultural Materials Initiative, Museum and Library Services, Council on Library and Information Resources[5]. This is fostering the reuse of cultural collections, through several media, for research and artistic purposes. The release under open licenses of the cultural collections also contributes to the introduction of cultural content aggregators, such as Wikimedia Commons and Europeana. They offer the possibility to access cultural objects across multiple websites.

Fig. 1.2
Johan Joseph Zoffany, The Tribuna of the Uffizi, 1772-77.

Opening a collection brings a set of pros and cons. On one hand, cultural institutions may have some concerns about the opening of their collections. These are mainly related to the complexity of legal and technological management. They also concern the fact that their physical presence may become obsolete. On the other hand, the main opportunities related brought by open licenses are the maximization of the visibility of the cultural institution and grants fundraising, the increasing of potential collaborations with other partners using open licenses and the cultural objects data enrichment. Furthermore, opening digitised collections, and gathering them in a unique, virtual space, can allow scholars to discover new insights (Gordon Lanning et al., 2016).
Nonprofit organisations are currently supporting cultural institutions in digitising and release under open license their collections through several initiatives. Between them, one of the most relevant is the development of the OpenGLAM principles[6]. They consist of a set of five essential guidelines that support the release of open contents over the whole range of the cultural institution activities, including the digitisation and user involvement.

1.1.1 Computerisation and digitisation in cultural institutions

In ancient times, cultural institutions used to keep track and manage their cultural objects through the use of paper records. Only in the 18th century, with the invention of the punched cards[7], a period of technological innovations aimed to automate some record-keeping activities began.
In the 1960s, the introduction of the modern computers brought new tools to access, process, and store information in a digital format. The early computers were large, expensive and demanded highly-qualified operators. They were first adopted by cultural institutions in the USA, in order to manage extensive collections. Several museums developed specific computer programs, such as GRIPHOS and SELGEM, and made them freely available for the museum community. GRIPHOS (General Retrieval and Information Processor for Humanities Oriented Studies), the earliest digital cataloguing systems, was released in 1967, and it is still in use in some museums (Kräutli, 2016). SELGEM (Self Generating Master), instead, was the most prominent software (Williams, 2010). Developed by the Smithsonian Institution, it consisted of a suite of programs created to manage the museum collection

based on the analogue and digital data entry technologies available at that time.
In the late 1960s, some standards for managing documents and other collection items were established. Among them, there were MARC (MAchine-Readable Cataloging), a set of digital formats for archival resources which nowadays is still the most used standard, and SGML (Standard General Markup Language), a markup language for documents.
In the 1980s, the computer desktop had a worldwide spread, thus every cultural institution entered the computer age (Williams, 2010). Images, videos and other multimedia files started to be gathered, stored and shared through local networks.
The digitisation became an essential activity of the cultural institution, because users demand for remote access and the majority of resources that are not electronically accessible are not used (Gwen, 2010). Furthermore, the digitisation has increased the visibility of the cultural institution.
The digitisation fulfils three central cultural needs of the institutions: foster the collection management, enlarge its audience and support the preservation of fragile and valuable artworks.
The digitisation process encompasses three main stages: digitisation, documentation and storage (Conway, 2014).
The digitisation phase involves the use of technological instruments, such as scanners and photo cameras, to digitalise the cultural objects (see fig. 1.3). Recently, tools like hyperspectral cameras and 3D scanners have provided new datasets for the research (Morgan, 2014). The documentation phase refers to the entry of data related to cultural objects. In this phase, machine learning technologies and crowdsourcing may be used to gather metadata. In the storage phase, digitised objects are stored on servers, with the aim to provide computer-readable texts and generate outcomes for the web (Bekiari and Constantopoulos, 2007).
The parameters of the digitisation process, such as format, size and specific metadata, are defined according to the expected use of the digital resource. The outcomes of the process must comply with interoperability requirements. In other words, they must be accessible through multiple information systems and durable for a particular time. (Panetto and Cecil, 2010)
The digitisation of cultural objects has changed many aspects concerning the supply and demand for cultural heritage. In particular, digitisation has generated a digital cultur-

al heritage which is no longer linked to a physical place and it has increased the consumer's expectations in terms of immediate availability of the collection and its participation in sharing the processes of knowledge (Benhamou et al., 2013). The digitisation process also involves several issues: the most relevant are related to operational costs, due to the shift from the analogue to the digital management systems (Coughlin et al., 2013; Gordon et al., 2016), the preservation of the digital archive (Meho and Tibbo, 2003), and the bias due to errors in the digitisation (Conway, 2014). Because of these issues, several cultural institutions have partially digitised their cultural objects or have not done it at all[8].

In the 1990s, the advent of the Web and the mass digitisation profoundly transformed the ways of documenting and managing the collection (Williams, 2010).

Cultural institutions realised that the online experience of their collections was complementary to the physical one. Thus, they adopted new tools to share information and multimedia content, such as newsletters and podcasts. They also implemented e-learning systems, digital archives, searchable databases and other online platforms, in order to improve the collection experience (Manovich, 1999).

In the late 1990s, the Web aspired to become a network of online platforms where contents were mutually shared. It was the so-called "web of data" or "semantic web". In order to formally represent the properties and relations between contents coming from multiple domains, the concept of ontology was introduced together with new standards (such as RDF) to exchange information among platforms. Nowadays,

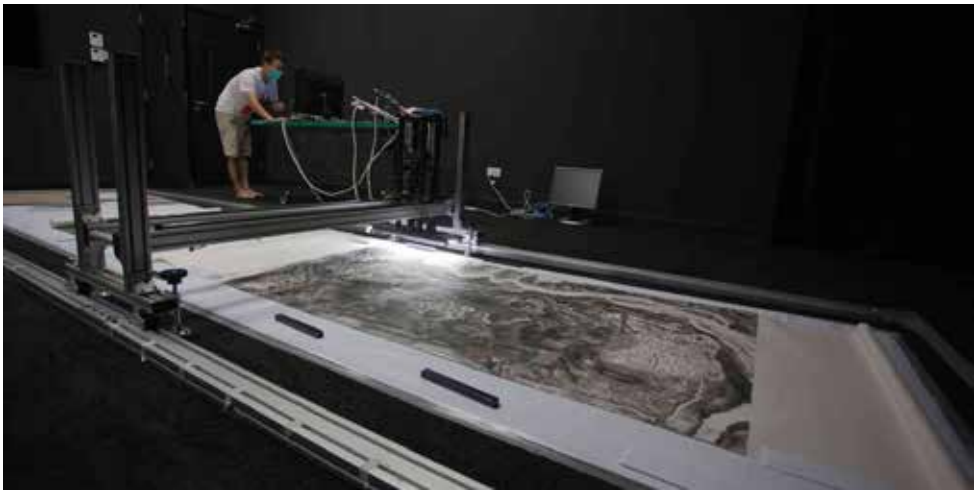


Fig. 1.3
Advanced analytic scanner.

a centralized system where contents float among multiple webpages exists and every cultural institution is free to adopt any standard.
Recent technological developments have favoured the access to a considerable number of cultural objects. The design of digital archives allows scholars and general audience to discover the cultural collection through intuitive and engaging user interfaces.

1.1.2 The introduction of open licenses

Digital technologies have deeply transformed the way cultural institutions fulfil their mission to preserve and provide access to the collections. Cultural institutions are primarily adopting digitisation technologies and web platforms to search collection databases. Both the digitisation and the use of online platforms raise opportunities and issues related to the copyright, a law which rules the ownership of the artworks created by an author[9]. It can be applied to pictorial, photographic, literary and any creative work made by humans.
The copyright law grants material and moral rights to the owner. As defined by the Berne Convention in 1979[10], material rights refer to the entitlement to reproduce the work, distribute copies, display it publicly and produce any derivative work. The authors - generally called "right holders" - hold the exclusive right to use or authorize others to use their work. The economic rights related to copyright begins with the creation of the work and last at least for 50 years after the creator's death (in most of the countries they last for 70 years). Moral rights - also called "ethical rights" - are generally considered inalienable[11] and thus they remain even if the copyright of a work is completely sold. The moral rights of an author include the right to claim the authorship of the work, the right to object any modification of the work, as well as any action that may damage its reputation.
In the European laws it is not required to register copyright in a designated office, nor it is necessary to include the copyright notice on a work in order to obtain copyright protection. In an original work, the copyright exists from the moment it is shown in a tangible medium. The benefits provided by the copyright are mainly directed to the author (or who owns the rights) and aim to foster the creativity and the dissemination of the work within a controlled context of use defined by cultural institutions, which can also earn

economic rewards through the management of the material rights concerning digitised artworks.

The concept of copyright was introduced by the English Parliament in the 17° century, after the invention of the printing press, as a way to prevent the publication of certain types of books.
The concept of open content was only introduced around 1940. At that time, several questions around the idea that information was based on a collective process arose and people were looking for strategies to generate knowledge through its free dissemination. (Wiener, 2014). The word "open" refers to the shared resources, accessible by everyone with no need of intervention from the author or any other right holders (Pomerantz and Peek, 2016). According to the Open Knowledge Foundation, "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)[12]".
The definition of open content is also related to the work of volunteers and people participation. In 1998, David A. Wiley introduced the Open Content Project, an initiative dedicated to generate works under a few usage restrictions (see fig. 1.4). It was designed for academics, but it was adapted to other scientific and artistic fields as well.

The Open Content Project was closed in December 2002 and it was followed by another one, the Creative Commons (CC),

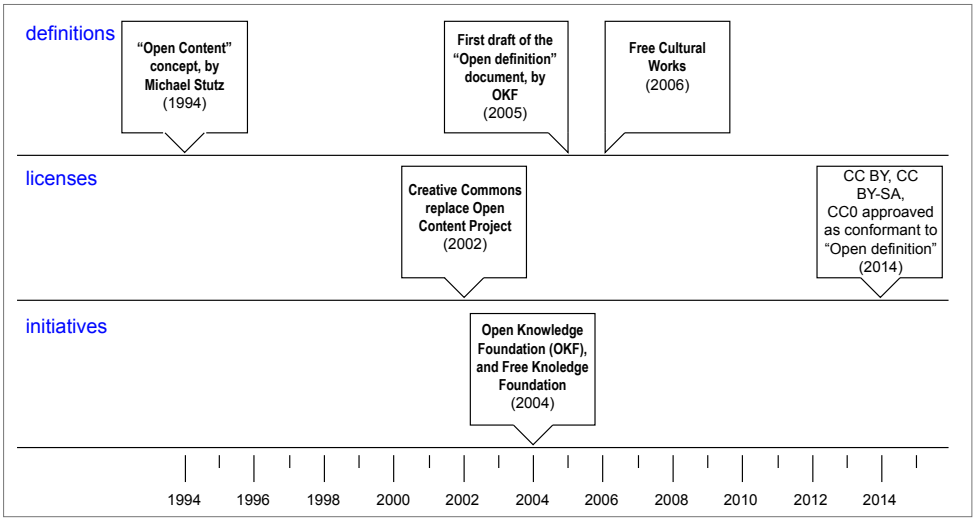


Fig. 1.4
Open culture timeline.

which was born thanks to an American non-profit organisation devoted to the release and sharing of creative works. The Creative Commons[13] project released several copyright-licenses, through which the author of creative works could decide to maintain some right. The Creative Commons licenses are defined by a declaration at the moment of the release and include the following conditions - clauses to be respected by law:

- PD (Public Domain), free usage of the artwork; everyone is allowed to copy, modify, distribute and perform the work, even for commercial purposes, without asking permission;
- BY (attribution), usage of the artwork with the author attribution;
- NC (Non-Commercial), usage of the artwork only for non-commercial purposes;
- SA (Share Alike), share of the artwork under identical terms;
- ND (Non-Derivatives), usage of the artwork as a whole, without derivatives.

These conditions can be combined (see fig. 1.5). Cultural institutions mainly release their digital collection under the attribution-share license (CC BY-SA) and under the Public Domain (PD)[14].
There is still a grey area on the rights which is arising from the digitization activity. On one hand, the most popular thesis is that the digitisation process does not produce any right on the digitised version of a cultural object while, on the other hand, someone supports the idea that it creates copyright due to the character recognition and the addition of metadata (Casella, 2013).
The benefits provided by the Creative Commons licenses are mainly directed to the community and aim to foster the production of collective knowledge through sharing and mixing practices.

Several cultural institutions are adopting Creative Commons licenses both for practical and prestige reasons. Practical reasons include the simplification of the terms of use, while prestige reasons are linked to the chance to obtain visibility and relevance and to avoid the online spread of fake artworks. Cultural institutions generally release only a small part of their digital surrogates under open license. In

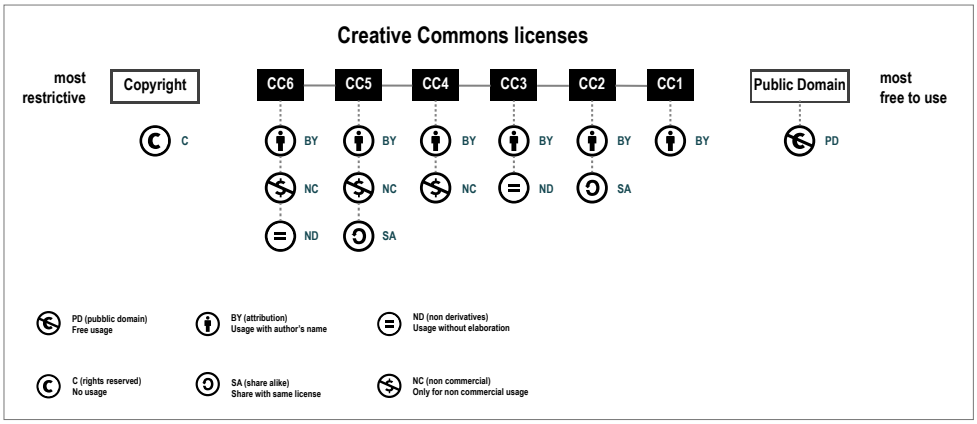


Fig. 1.5
Creative Commons licenses.

this way, they maintain the majority of the revenues coming from the commercial use of their digitised collections (such as reproductions, gadgets, publications, etc.).
Currently, there are conflicting opinions about whether to use or not open licenses for digitised cultural collection. The main worries concern the possible loss of revenues from rights and reproduction of digitised materials and the loss of intellectual control (Kapsalis, 2016). Furthermore, cultural institutions fear potential losses of attribution and the illicit use of cultural resources. The digitisation, cataloguing and preservation activities are expensive. For this reason, cultural institutions are hesitant to release digitised collections under open license.
On the other hand, their main advantages are related to an increasing in terms of visibility and grants from fundraising, as well as possible collaboration with other partners using open licenses. Through the release of the collections, cultural institutions increase their online discoverability and improve the opportunities for their audiences to participate in the curation and enrichment of their collections.
Some cultural institutions, such as the Rijksmuseum, release their digitised collections at a medium resolution and under open licenses, but also provide high resolution access after the payment of a small fee.
The opening of cultural collections has several advantages for the society as well. Sharing open cultural objects and their related information allows scholars to study artworks and people in a more comprehensive way, enables teachers to reuse the artworks in the classroom and give the citizens the chance to enjoy, understand and reuse their cultural heritage in order to produce new artworks.

1.1.3 Reuse of digitised collections

Who produces a digitised copy of an artwork, by law, owns the copyright. The related metadata, instead, do not provide rights. The author owns the rights derived from the digital copy of the works. Currently, several cultural institutions have realised their digitised collections under open licenses (mostly) or in public domain. Among the most relevant ones, there are MoMA, Smithsonian Design Museum, New York Public Library and Rijksmuseum (Netherlands).
The release of digitised collections is fostering their use not only by scholars, for research and exploration purposes, but also by organisations and ordinary people for the reproduction on digital and paper supports. An example is "The Public Domain Review"[15], an online journal founded in 2011 and realised by a community of writers, which explores works of history, art and literature that are in the public domain. The journal aims to disseminate knowledge and promote the public domain in all its abundance and variety.
In addition to traditional uses, new, unconventional practices - such as the sharing activities and the remix- are emerging, redefining the way collections are used.
Sharing is mainly related to social networks, where end-users share digitised objects with their followers just to provide them some inspiration and to represent current status. On the other hand, the remix refers to the graphical elaboration of one or more digital copies, in order to give them a new meaning and value. The "remix culture" dominates the 2000s, it affects multiple cultural sectors and it is governed by fusions, collages and mashups (Manovich, 2007).
The reuse of digitised cultural collections has essentially an artistic and ludic purpose. An example of the artistic use of digitised collections is the project Succession[16] by Mitchell Whitelaw, which automatically generates a composition starting from images documenting the history of Newcastle-Upon-Tyne. Each one is a unique combination of five random items from a set of about two thousand images.
An example of ludic use of digitised collections is "GIF IT UP"[17], promoted by Europeana, which is an annual international contest encouraging the content aggregator community to generate unique and fun gif artworks starting from digitised cultural heritage resources. The users are invited to search, discover, adapt, and reuse public domain images found on cultural content aggregators which are international partners of Europeana, such as Trove[18] and DigitalNZ[19].

1.1.4 The open cultural content ecosystem

Following the opening of digitised cultural heritage, the landscape of actors producing and sharing digital contents is getting wider and wider. Every actor aims to maximize the quality and make the offered content or service gain a high online visibility. They are moved by several reasons, ranging from the pure spirit of activism and volunteering to commercial purposes. As though in a natural ecosystem, every actor interacts with the others and with their artefacts and services. Within the ecosystem, the single digitised object acquires a new meaning through its remix and re-contextualisation (see fig. 1.6).
The following paragraphs explain the figures and logic behind what we can define "the open cultural content ecosystem".

Cultural institutions
Cultural institutions are galleries, libraries, archives, Museums (GLAMs). They play a crucial role by defining the requirements an item must fulfill to be digitised and released as an open content.
Usually, a GLAM digitises its entire collection in a systematic way or upon a request made by scholars, curators or patrons. The digitisation can also take place before or after events like cleaning and repairing processes.
The choice to release specific digitised objects under open

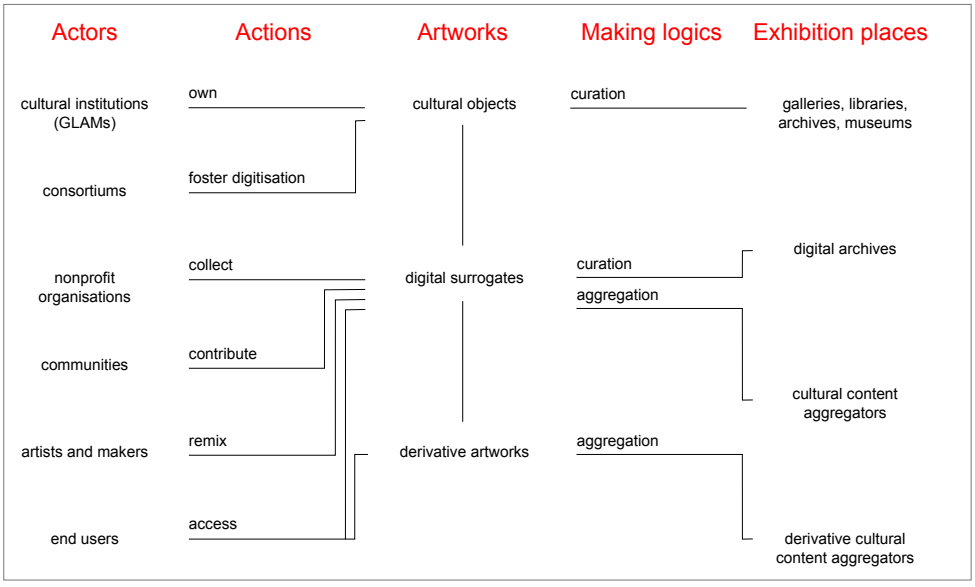


Fig. 1.6
Open cultural content ecosystem.

licenses is based on two main approaches. First, the GLAM releases a set of digitised objects as a sample of the whole collection and then it releases an entire set of digitised objects belonging to a specific theme.
GLAMs release the digital surrogates of their cultural objects through their web platforms, specifically in their digital archive or their API (a set of functions to access contents programmatically), and cultural content aggregators. The content is then promoted via the institutional website, press releases and social networks.
The figures who have key roles in the opening of the digitised collection are the digital collection curator, the responsible for digitisation and the community manager[20].
The digital collection curator is responsible for the selection of cultural objects to be digitised and for the managing of their digital version. Due to the digital obsolescence[21], he has to perform several actions in order to keep the digital assets and their related metadata readable over time.
The responsible for digitisation plans and supervises digitisation process of the collection. In agreement with the digital curator and in compliance with the technical feasibility, he defines the goals and priorities of the GLAM digital strategy.
The community manager supervises the relationships between the GLAM and its community. With regards to the collection, he gathers the needs of the community concerning the digitised items which are not available yet and fosters the use of the digitised collection through training courses, workshops and hackathons.
Among the most relevant cultural institutions in Europe there are the Rijksmuseum (in the Netherlands)[22], The British Library[23] and the Deutsche Nationalbibliothek[24].

Digitisation service providers
Digitisation service providers take care of the digitisation of the GLAM cultural collection. They choose the most suitable technologies and formats in order to satisfy requirements of the cultural institution.
Digitisation service providers are fundamental not only because they foster the online access of digitised artefacts, but also because they contribute to the preservation of the original artworks. Digital copies allow cultural institutions to protect the original ones from handling, while presenting their content to a vast audience.
Until ten years ago, the digitisation process was performed

by the GLAM itself or by public digitisation centres only. Today, due to the continuous update of the minimal digitisation standards of the processes and their related, considerable costs, several big European and American private companies (Google and Microsoft among them) are entering the cultural content system by starting mass digitisation projects (Casella, 2013). Among these projects, there are Google Books, Google Arts & Culture[25], Microsoft In Culture[26] and Hathi Trust[27].

Nonprofit organisations

Nonprofit organisations promote the opening of digitised collections and their spread. They lead the development and the management of cultural content aggregators which share the same subjects or geographic areas. Nonprofit organisations also boost the relationships between GLAMs and communities by organising workshops and contests. Moreover, they can obtain financial resources from the public and private sectors.

As pressure grows for cultural institutions to provide open access to digital content of collection objects, traditional models of image reproductions are being re-evaluated and adapted. Provoked by the revolution affecting how people access and use digital media, organisations are leading the sector through this cultural shift, by providing education, guidance and new types of image licence agreements (Powel, 2016). Among these organisations, there are Open Knowledge International, OpenGLAM, and Free Knowledge Foundation[28].

Among the most known nonprofit organisations, there are the Open Knowledge Foundation[29], the Wikimedia Foundation[30] and the Europeana Foundation[31].

Consortiums

Consortiums are aggregations of GLAMs, nonprofit organisations and individuals devoted to support the digitisation and sharing of cultural objects under an open license. These networks also foster the sharing of ideas and best practices concerning preservation and communication of cultural heritage. Sometimes, a consortium consists of partners sharing the same country and regulations[32], such as the GLAM-Wiki US Consortium, a network of US GLAMs collaborating with Wikimedia Foundation.

A consortium can also be international and consist of GLAMs sharing collections with the same subject or medium. An ex-

ample of this kind of consortium is the Photoconsortium, a network for the photographic heritage, based in Italy, which provides tools and methodologies for the digitisation and managing of photographic images.

Communities

Communities voluntarily contribute to expand and enhance digital collections released by GLAMs.

In recent years, bottom-up communities of individuals who self-organise to produce goods, services and contents are spreading worldwide. These community-based organisational forms are flexible and highly scalable: the more participants they are, the earlier they reach the goal. The structure of a community is a network taking shape accordingly with the outcome to be obtained. The early communities were born to solve local problems sustainably (Menichinelli, 2009). A community can be self-managed or managed by a nonprofit organisation or a company. Thanks to the use of web technologies, a community can connect people coming from all around the world in order to generate artefacts and contents with a high level of complexity.

According to the cultural aggregator features, communities can organise their contribution activities both in a hierarchical or not hierarchical model. In a hierarchical model, an editor supervises complex contributor actions, while inn a not hierarchical one each contributor is able to perform simple actions. Among the biggest and most popular communities, there are Wikipedia and Europeana communities.

Communities producing content strongly rely on the concept of "open content" as an easy way to reuse works by many individuals following the copyright laws.

In the context of cultural collections, the work of a community is particularly relevant because, through the reuse of digitised cultural objects outside of the cultural institution website, it enhances their visibility and perceived value.

Artists and makers

Artists and makers are people who use digitised cultural artefacts for creative purposes and to give them an added value. Artists use digital technologies both to perform their creative actions (technology as a tool) and to make generative works of art (technology as a generator).

Makers are intended as passionate people using digital technologies for practical purposes. They contribute to the diffusion of the digitised collection through the making of

derivative works and the development of information systems that aims to enhance the accessibility and the quality of open cultural content. An example of the derivative cultural content aggregator is What's the Picture[33], an online tool by James Marley which allows the enhancement and enrichment of images coming from cultural collections and uploaded on Flickr under Creative Commons licenses.

End-users
End-users are mostly students, scholars and writers (including journalists and bloggers) using digitised artworks for their study and dissemination activities.
In contrast with digital archives users, aggregators end-users have a research theme; precise ideas, typologies and characteristics of images to be investigated. In some cases, professional photographers working for communication agencies collaborate with writers and publishers to select the digitised artworks to be included in publications. Papers, articles and online portals are some of the outcomes where digital cultural objects are added.

1.1.5 Rise of cultural content aggregators

Cultural content aggregators aim to gather open multimedia contents (such as images, videos and sounds) coming from several sources, in order to foster the user's access and use. Most of cultural aggregators are based on the contribution given by verified GLAMs. Communities and end-users can contribute to expand and enhance the contents via online forms or through the backend user interface after completing the registration procedure. Aggregators can allow a wide range of user interventions, according to the terms of use and the goals of the organisation which manages the platform.
Recently, generic content aggregators based on users' contributions are partially acquiring the role of cultural aggregators. Among them there are Pinterest and, in particular, Flickr, which has launched The Commons[34] in 2008 (see fig. 1.7), a section where users can contribute with photographic images coming from worldwide public archives under Creative Commons license. In this kind of aggregators, user-friendly interfaces allow users to upload and access digital objects.
Generic content aggregators were born in the 1990s, when there was an emerging need to centralize multimedia con-

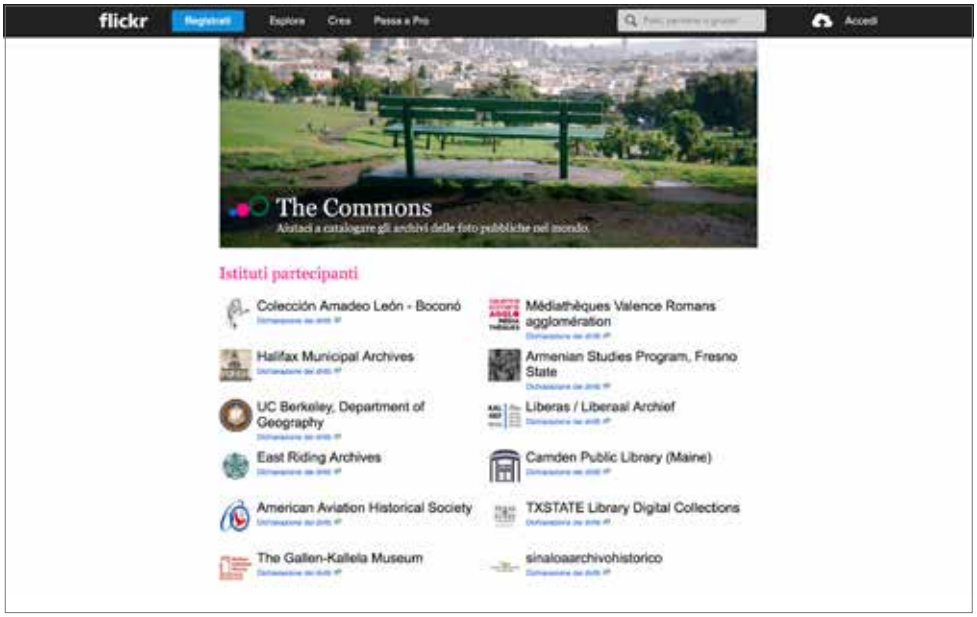


Fig. 1.7
Flickr The Commons.

tents from several sources to a few, big platforms. At that time, when research engines were introduced, content aggregators offered the possibility to access contents through multiple websites.

One of the very first cultural content aggregators was WorldCat[35]: launched in 1998, it is a still active catalogue of items coming from more than 70.000 libraries in 170 countries. WorldCat does not provide information on the current state of the items (including loans and undergoing restoration processes) but it links the item to the institution which owns it.

In the 2000s, the current, biggest cultural content aggregators, such as Wikimedia Commons (2004), The European Library (2005) and Europeana (2008) were launched.

Cultural content aggregators index the contents through a researchable database, in order to make them easy to retrieve. In some cases, a cultural aggregator consists or it is provided with a web platform (called wiki) which allows users to add and edit contents.

The introduction of cultural content aggregators led to a shift in the paradigm of accessing digitised cultural content. While digital archives, containing a single collection, consider curation as an approach aimed to select and organise the contents, cultural content aggregators gather contents from multiple archives through aggregation.

The curation is performed by the digital curator on behalf

of the cultural institution: it consists of an accurate work of selecting, adding value and presenting the digitised collection through a virtual space, such as a digital archive, official institution website and mobile application. The valorisation of the single digitised object can occur through the addition of related information and content, while end-users are usually guided to a predetermined fruition path. Due to the hand-picked digitised objects, curation is a time-consuming process, but it focuses on the quality of the content.
On the other hand, aggregation is a process which requires the contribution of multiple individuals, aimed to gather digitised cultural objects coming from different contexts and to show them as a stream of contents. In this case, end-users are more active than in a digital archive, because he requires specific digitised objects via specific interface tools. The added value of the aggregation is the possibility to access items coming from many sources through a single web platform. Moreover, it uses a rapid process, where the quality of items and their related information can vary.
Despite these differences between archives and content aggregators, they have two relevant aspects in common: the connection of end-users and the user-generated content. The first one refers to the fact that both web platforms foster communication among cultural institutions and users, while the user-generated content allows users to contribute to their content.

1.2 Classifying and managing open collections

Archival science is the discipline that categorises and manages cultural collections. Archival science, or archival studies, theorizes the building and curating of collections of documents and data (Pearce-Moses, 2005). It aims to improve methods to store, classify retrieve and preserve cultural objects.
Classification and preservation systems were applied to archive items since ancient times, but archival studies were recognized as a research discipline, coming from diplomatics[36], only in the 16th century (Pearce-Moses, 2005). The advent of the digitised cultural objects and the development of electronic databases has caused the re-evaluation of this research discipline: the need to enhance the discoverability of archival materials, both for experienced users and beginners, led to manage sets of metadata consistently and

to develop flexible classification systems. In the academic field, archival science is conceived as a branch of information science[37].

Cultural analytics is another discipline which studies cultural collections is. It analyses massive cultural data sets as visual materials. Even though this thesis does not focus on the visual analysis of cultural collections, cultural analytics is relevant because of its use of graphical user interfaces to explore extensive collections (Manovich, 2012).
In cultural analytics, the objects of the study are both digitised visual artefacts and contemporary born-digital artworks. The term "cultural analytics" was coined by Lev Manovich in 2007[38].
Cultural analytics uses visual techniques based on computation and image processing to explore extensive collections. It also develops a graphical interface to improve the discovery of visual patterns by researchers. Among the methods for the analysis of visual cultural artefacts, there are the visual mapping of multiple surrogates according to specific visual variables, the data mining and data visualisation of metadata. With regards to the analysis of bibliographic works, the interest for the distant reading techniques is increasing. Distant reading is a professional reading methodology based on computer programs (Moretti, 2013), which can be considered as an attempt to adopt big data analytics methods for literary scholarship.
Cultural analytics is a discipline under the umbrella of the Digital Humanities (DH), a research area at the intersection of computing and the disciplines of the humanities, which mainly uses digital resources and a transdisciplinary approach.

The following chapters present the main concepts and topics of archival science in the context of open cultural collections. They also try to reframe some content according to the recent developments in cultural content aggregators.

1.2.1 A taxonomy for the digitised collection

In cultural analytics, research focuses on visual cultural objects and their digitised versions. The study of cultural patterns throughout history and nowadays requires, on one hand, the study of requirements and characteristics of cultural objects and, on the other hand, an analysis of the visual features of their related digital surrogates.

This research thesis focuses only on the digital surrogates coming from cultural artefacts and, specifically, on digital images. Usually, they are obtained through the photograph or the scanning of the cultural objects. Images can also be generated in set (like in the panoramic photography or in the 3D scanning) and be further elaborated in order to become a machine-readable text (like in the Optical Character Recognition OCR). For this work, we will only consider images coming from visual artworks (in 2D), such as photos, paintings and prints.
In recent years, digital images are becoming a relevant content on the Web because of their ability to communicate information and emotions immediately. In order to better understand the origins and the characteristics of the digitised cultural objects, the single aspects of the digitised collection are presented.

Cultural collection
A cultural collection consists of a set of artefacts related to each other by some common categories determined by human beings (Currall et al., 2005) or physically arranged together (Zeng and Qin, 2008). The categories can refer to the origin of the artefact, such as subject, author and medium. Generally, artefacts are collected over time through different acquisitions modes: donations, inheritances, acquisitions and field research (Marini Clarelli, 2005, p. 12).
Cultural collections play an important role in witnessing knowledge, habits and feelings about an author, an historical period or a topic. The key factors of a cultural collection include selectivity, flexibility and "ready availability". Selectivity refers to the fact that the collection is curated, flexibility means that a collection can be shaped according to the user's needs and the "ready availability" intends that only ready, available items can be considered as a part of a collection (Lee, 2005). In the past, many cultural objects have been moved from their place of origin. For instance, the Parthenon marbles are in the British Museum. Nowadays, responsibles of collections remove artworks from the context of origin only if it necessary for conservation and safety reasons (Marini Clarelli, 2005, p. 12).

Cultural object
The term "cultural object" refer to an artefact belonging to a cultural collection. A cultural object is an original, tangible expression of the idea of a human being (Wallace and Ronan, 2016) within a specific cultural context (Bearman and Trant, 2008). It can be either physical or digital. Physical

cultural objects include paintings, photographs, sculptures and manuscripts. Digital cultural objects are born-digital objects and include digital paintings and videos, software and screen-based applications. Cultural objects are relevant because they testify several aspects of the society, such as the technological advancements, the economic development and the social hierarchy.

Digitisation

Digitisation aims to make a digital representation of a physical cultural object available. The representation does not replace the cultural objects but it provides additional information layers that enhance its comprehension (Stiller, 2012). Digitisation must be distinguished from conversion: the term digitisation refers to all the processes that change the meaning and value of cultural objects, such as scanning a photograph, while the term conversion is used for processes that maintain the cultural value of the original object, such as the digitisation of a poem (Benhamou et al., 2013). The digitisation of cultural objects is considered to separate their form from their content (Burns, 2017).

The digitisation process, also called digitisation program, consists of three main stages: the selection of the cultural object to be digitised, the digitisation through photography or scanning[39] and the generation of metadata describing both the cultural object and its digital version.

There are two digitisation outcomes: the digital surrogate and the digital addition. The digital surrogate refers to a digital item representing the physical cultural object completely, acting as a substitute, while the digital addition (or metadata) concerns the digitisation of the data related to the cultural object (Stiller, 2012), and can be generated both by humans and computers.

A digitisation program creates a digital artefact that has the technical characteristics and the traces of the time when it has been produced. The derivative artefact may contain errors related to the digitisation methods adopted (see fig. 1.8). A digitisation program should include not only the digital capture of cultural objects but also an appropriate long-term management of the digital files it produces (Matusiak and Johnston, 2010).

Digital surrogate

The reproduction of a physical cultural object produces surrogates. These entities are tangible representations that

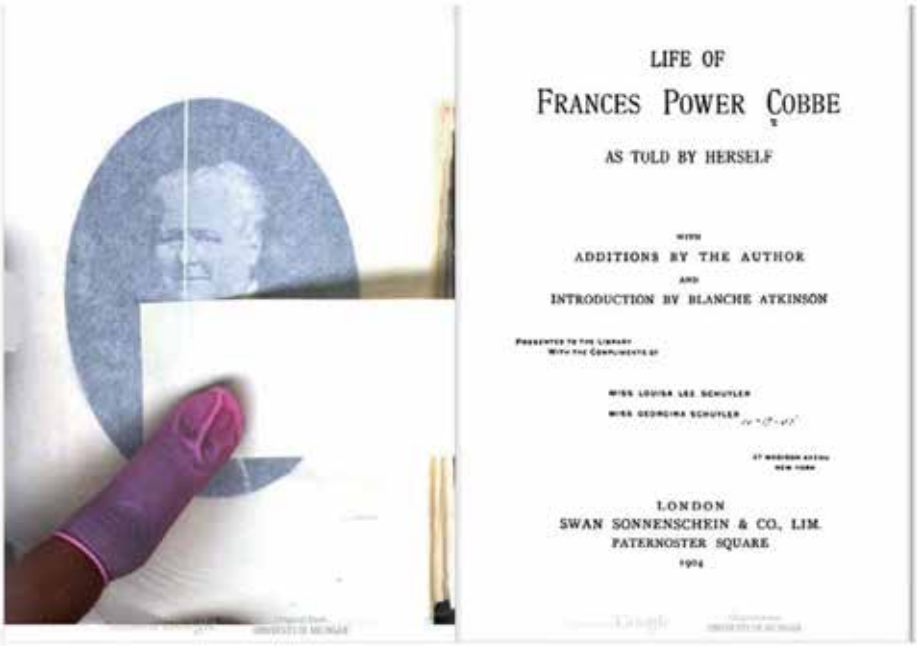


Fig. 2 Finger, tab, tissue, book illustration (Cobbe and Atkinson 1904, p. ii)

help to capture the memory of the cultural objects and to preserve it for future generations. The production of surrogates (or proxies) considers two main aspects: identity and fidelity. Identity refers to the goals of the digitisation and answers the question: What is the aim of the surrogate? Fidelity, on the other hand, concerns the similarity between the objects and the surrogate. It answers the question: How similar is the surrogate in comparison with the object? (Davis, 1995). A surrogate can be both physical or digital, and it is as tangible as its analogue counterpart (McCarty and Kirschenbaum, 2003). Physical surrogates include photographs and 3d printed models. While digital surrogates include digital images, videos and 3D models.

Even though a digital surrogate represents a cultural object, it has a different status. While the cultural object has a unique existence in time and space (the *hic et nunc* of the work of art), the digital surrogate has multiple copies that can be watched over time and in several locations. The digital surrogate lacks the "aura" of the cultural object (Benjamin, 1936). It also lacks or does not perfectly represent some of the characteristics of the original one, such as textures,

Fig. 1.8
Errors in digitisation, The art of Google Books, by Krissy Wilson. Website: http://theartofgooglebooks.tumblr.com/

sounds and smells. Despite these differences, researchers are more likely to study using the digital surrogate than the original cultural object. In most of the cases, the original is owned by a private and the digital surrogate is the only way to access and study it (Shein and Lapworth, 2016).

Digitised cultural collection

A digitised cultural collection is a set of digital surrogates belonging to a cultural institution. The digitised cultural collection has three main goals: provide access to the collection without restrictions of time and space, preserve the physical cultural objects and enhance the value of the cultural institution.

Nowadays, digitised cultural collections are particularly relevant in creating novel forms of interaction with online and on-site visitors (Monaci, 2005, p. 27). The main channel to access digitised cultural collection is the Web. Its properties allow users to obtain rich overviews of the collection, high-quality digital surrogates and detailed information.

Metadata

A digitisation process generates a digital surrogate, as well as metadata. Metadata are related both to the cultural object and the digital surrogate.

The term metadata was introduced in the English language in 1968. It stands for something that is beyond the data, at a higher level of abstraction. They can be seen as a map through which it is possible to represent an object in a more straightforward form. In the cultural context, metadata refers to objective, structured information about a resource of any media type or format (Stiller, 2012). The metadata is mainly thought of being shared through an information system and its use started centuries ago, when librarians adopted it as a way to help users to find materials in the library's collection (Caplan, 2003).

A single metadata consists of a couple of items or statements: a property defining the meaning of the data and a value for example, author (property) - Leonardo da Vinci (value) (Zeng and Qin, 2008).

The metadata is not merely a data about an object (either physical or digital), but it refers to an object in a particular context, generated by a particular individual or organisation. Since organisations differ in outlook, capabilities, and audiences served, the metadata produced will necessarily reflect those different contexts (Burdick, 2012).

The most influential institutions collaborating to define metadata standards for cultural objects and related digital surrogates were the Getty Information Institute and the Visual Resources Association (VRA). In the 1990s, the Getty Information Institute published several standards that nowadays are continuously updated. The Visual Resources Associations still contributes to spread metadata best practices within museums, libraries and archives (Pomerantz, 2015).
The metadata consists of four typologies: descriptive, structural, administrative and usage metadata. Every metadata typology has its own rules defining how to register data (Pomerantz, 2015).
Descriptive metadata concerns the cultural object. Among these data, there is the location of origin, the author, date and keywords. These data allow the identification of an item; thus, it is an essential means to access the cultural objects within a collection (Caplan, 2003).
Structural metadata refers to data about the object structure and how the components of an object are organized. Among this data, there are dimensions, materials and colours (Pomerantz, 2015).
Administrative metadata are about the provenance and the maintenance of the cultural object, such as the categories, the location and the license. Furthermore, administrative metadata include data about the classification of the digital surrogate and its relationships with other surrogates (Pomerantz, 2015).
Usage metadata (or paradata) concerns the usage of the cultural object and its digital surrogate. In technical settings they are called logs (Pomerantz, 2015). Among the usage metadata there are loans, downloads, views and likes. Currently, there is a growing interest in this kind of data because they can reveal insights about the fruition patterns of cultural resources.

Metadata set
The metadata set is to be used for every single item. The most common metadata set includes title, author, tags (or classes), typology (or medium), description and the GLAM owner. Many metadata set also include elements which indicate the relationships with related cultural objects (Pomerantz, 2015). In order to easily retrieve and share digital surrogates, cultural institutions should collect various types of metadata following standard rules. Due to the rapid growth in digital object repositories and the development

of many different metadata standards, metadata generation is becoming more and more complex (Sotirova et al., 2012).

Class
The class is a set of digital surrogates with a metadata in common. The action of placing digital surrogates into classes according to their metadata is called "classification", or "categorisation". Since every information system strongly rely on the classification, the selection of classes is crucial in the retrieval of digital surrogates.

Classification schema
Classification schema refers to the visual representation of the relationships sets of digital surrogates having common metadata, or classes (see fig. 1.9). Classification schemas include lines, networks, trees and clusters and can represent classes belonging to hierarchies or simply connected because of some similarities.

1.2.2 Typologies of classification systems

Classification is one of the most basic and universal human activities. It is a process that leads people to distinguish and organise everything surrounding them; in other words, classification is how humans bring order to the world (Roberts, 1997). Classification is applied in several thematic fields (such as science, literature and management) with different degrees of extensions: we can have sectoral (or disciplinary) classifications, such as the Standard Industrial Classification (SIC);

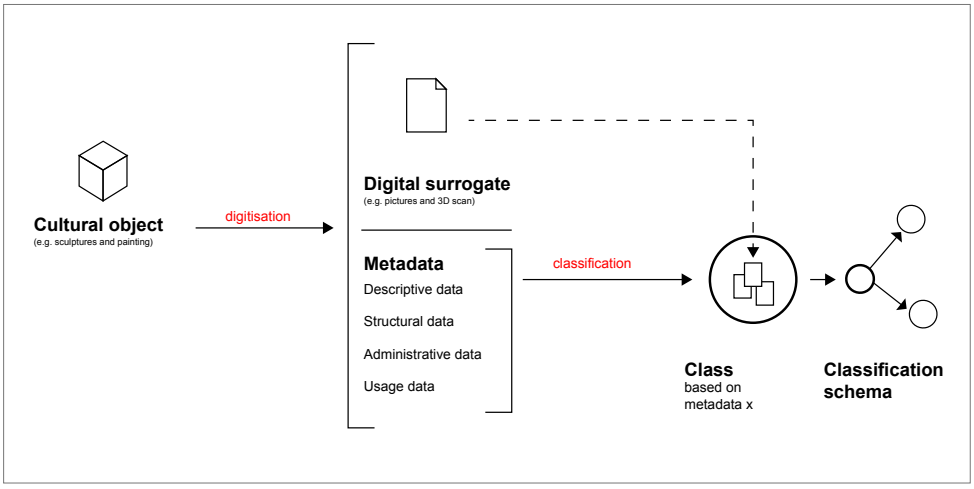


Fig. 1.9
Digitisation: from cultural object to digital surrogate and metadata. From metadata to classification systems.

national classifications, such as the Swedish library classification system[40], and universal (or encyclopaedic) classifications, such as the Dewey Decimal Classification.
Classification systems gather items according to common relations or affinities. They allow to quickly find items, identify any duplicates, understand the context surrounding an object and give it a meaning.
Every classification system requires the creation of metadata and the use of a specific type of notation (generally it consists of a set of numbers and letters). Due to the evolution of human needs and discoveries, classification systems must be frequently reviewed, entirely or partially.
In the context of cultural collections, classification systems are tools that facilitate access to knowledge and artefacts.

The Pinakes was the first classification system in history. Developed by Callimachus, a scholar working at the Library of Alexandria during the third century BCE, the Pinakes consists of a set of tablets that groups by subject the nearly 500.000 papyrus scrolls contained in the library at that time. Within the Pinakes system, works were listed by title, author and genre (Bagnall, 2002).
For centuries, libraries have used "shelf lists", catalogues gathering bibliographic works according to the acquisition order and the disposal on shelves. Then, in 1780, card catalogues were introduced in France. Thanks to them, every entry started to be managed independently from the others and they also started to be classified both according to its subject and its author (Pomerantz, 2015).
Classification systems in libraries, in general, have two main functions: facilitating the access to the item by allowing users to find out a subject, as well as providing information about its physical location. The most used library classification systems are the following: Dewey Decimal Classification (1876), Library of Congress Classification (1897), Universal Decimal Classification (1905) and Colon Classification (1933).
The Dewey Decimal Classification (DCC) was first introduced in the United States: it uses a relative location and a relative index in order to allow new books to be easily added in their appropriate location according to their subject. It consists of ten classes, each divided into ten divisions having ten sections each. DCC uses Arabic numbers as a notation and it organises the resources by discipline.
The Library of Congress Classification (LCC) is mainly used

in U.S. libraries. It uses the concept of relative location and divides subjects into broad but not universal categories. LCC consists of 21 classes, each divided into a number of sub-classes. The system uses alphabetic notations.
The Universal Decimal Classification is an arrangement of all branches of human knowledge. It is managed by the UDC Consortium, a non-profit international organisation based in the Netherlands, and applied to libraries worldwide. UDC uses ten classes, each divided into auxiliary sub-classes, and Arabic numerals arranged decimally as notations.
The Colon Classification, developed by the mathematician S. R. Ranganathan in 1933, is especially used in libraries in India. It was the first using a faceted (or analytical-synthetic) classification system consisting of a set of facets which describe the features of the bibliographic resource. CC uses 25 classes, each divided into some sub-classes, and five primary facets (or categories): subject specificity, subject property, processes, space and time. CC uses both alphabetic and punctuation notations.
Generally, classification systems which use multiple classification levels are more suitable to facilitate the access to the subject, but they are challenging to use for shelf arrangement (both in a physical and digital setting). Conversely, more flat classification systems make the shelf arrangement easier, but do not adequately show relationships between subjects.

Museums, archives and galleries have traditionally focused more on the documentation than on cataloguing their cultural collections (Zoller and DeMarsh, 2013). Within these cultural institutions, classification systems have followed the provenance principle for centuries. It refers to the classification of items according to the author and the order of creation. After the French Revolution in 1789, there was a transition period where multiple classification systems were applied and the concept of preservation arose. In the mid-nineteenth century, the provenance principle has been resumed because of the increasing number of archive objects. Recently, museums and art galleries are organising more and more thematically (Szostak, 2016).

Classification systems can be described according to the typology, the mode of use and the schema. The typology refers to the name with which a classification system is generally recognised, the mode of use concerns the arrangement of the items, and the schema refers to the graphical rep-

resentation of the relationships among classes.
There are three typologies of classification systems: nomenclature, taxonomy and faceted (or analytical-synthetic).

Nomenclature
Nomenclature is a collection of items or concepts in alphabetical order which focuses on the relevance of the item and aims to define in a simple, stable and universally accepted way, a list of terms or names as a basis for indexing and cataloguing human-made objects (Satija and Martínez-Ávila, 2015). In a nomenclature, successive divisions of a class can only cover one type of relationship adequately (such as hierarchical) and this may lead to unnecessary repetitions of topics (Buchannan, 1979).
With regards to the mode of use, nomenclature is enumerative. The term refers to the alphabetical ordering of all the classes, that are enumerated according to specific characteristics. In the nomenclature, the classification schema is linear: every class (or group, or node) must contain unique items.
Due to the reduced flexibility of nomenclatures, nowadays they are a bit obsolete and other classification systems are adopted (Satija and Martínez-Ávila, 2015)

Taxonomy
Taxonomy is a hierarchical list of items or concepts and focuses on the relationships between items, aiming to define hierarchies that could be manageable over time.
In terms of usage, taxonomy divides the subjects hierarchically, from the most general to the most specific one. In the taxonomy, the classification schema is a tree: a class (or node) belongs to a parent class and may have its own sub-classes (Martínez-Ávila, 2015). Every class must contain unique items.
Modern taxonomies, in particular in the scientific field, has been described as "basically a Renaissance codification of folk taxonomic principles". The expression folk taxonomy refers to the way rural people use the language to organise the objects around them (Raven et al., 1971).

Faceted (or analytical-synthetic)
This kind of classification system consists of multiple classes assigned to the same item. Classes can also be referred to as categories or tags.
These classifications aim to make items accessible through

the list of their various facets or properties. This approach avoids the need of a long list of classes and reduces the complexity of the notations (Martínez-Ávila, 2015). In faceted classification systems, the classification schema is the cluster and every cluster contains facets. A facet shares items with other facets accordingly to their properties. In some cases, the classification schema consists of a graph where facets are directly connected or nested inside each other. In the past, the most well-known analytical-synthetic classification system was the thesaurus, vast collections of words related to a certain period or sector. Nowadays, faceted classification systems are widely adopted by physical and digital archives.

Ontology
In Information Science, the ontology is a classification system where nomenclature, taxonomy and facets work together to organise a subject area of human knowledge. Within ontology, items, classes, attributes and axioms are interlinked with each other. The ontology classification schema is both tree and graph (Fu et al., 2013). An example of ontology is the one used by Wikidata, a collaborative knowledge-base managed by the Wikimedia Foundation. It consists of items, which represent topics, concepts, or objects, that are identified by a unique number over multiple language versions.

In a classification system, different usages can coexist. For instance, some of the most common library classification systems (Library of Congress Classification and Dewey Decimal Classification) are essentially enumerative, but with some hierarchical and faceted elements (Kumar, 1991). There are also online classification systems using multiples typologies of classification, for example the face-tagging (a combination of faceting and tagging), used in e-commerce, and the tags-taxonomy, used as the outcome of artificial intelligence processes.

1.2.3 Surrogate and metadata management

In general terms, surrogate and metadata management consists of all the activities related to the preservation and access of the cultural object over time. This is one of the primary responsibilities of cultural institutions (Stiller, 2012). In the case of open digital surrogates, new management aspects are arising. On one hand, there is the digital preser-

vation and, on the other hand, the monitoring of the usage. Digital preservation refers to the activities that the GLAM carry out to avoid digital obsolescence of the surrogate and to keep their metadata updated. Usually, cultural institutions which own digital surrogates set a digital preservation plan to assess their records, in order to identify any risk of file inaccessibility. In this sense, the Digital Preservation Coalition, a not-for-profit company based in the UK, constantly monitoring technologies and typology content that risk to become obsolescent (see fig. 1.10).

Among the strategies aimed to avoid file obsolescence, there is the use of open standards and formats (which allow the reuse of the source code from other hardware), the copy of the surrogate in analogue storage systems, the print and storage of physical copies and the conversion of the surrogate format into current file formats. Different formats may

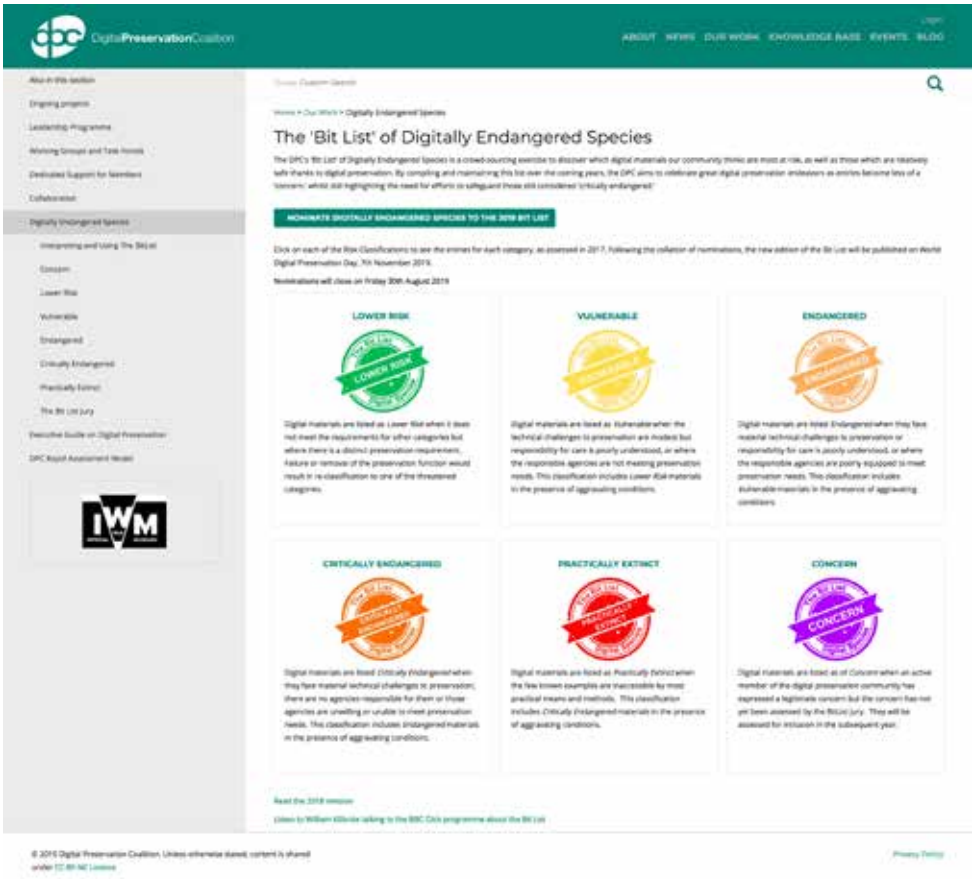


Fig. 1.10
The 'Bit List' of Digitally Endangered Species. Website: https://dpconline.org/our-work/bit-list.

need different preservation strategies and demand different technical solutions.
Within the digital preservation workflow, there is an increasing interest in the generation of metadata via automatic systems, such as support vector machines[41] and machine learning algorithms (Hodge, 2000).

When digital surrogates and their related metadata are released under open licenses, there is a need for the GLAM to maintain the quality of the records while uploading in a cultural content aggregator and to monitor the usage. Some cultural content aggregators make available uploading tools that save images and metadata in batch. An example is the Upload Wizard by Wikimedia Commons, a tool that helps to upload multiple multimedia files in four simple steps.
Europeana adopts a different approach. It collects digital surrogates coming from multiple cultural institutions through the Europeana Media Proxy (EMP). It is a software that gathers digital surrogates and metadata from the URL address provided by the institution partner and displays them through a standardised viewer. This information system makes it easy to maintain the consistency of metadata over the two web platforms.
Other increasingly relevant instruments for managing a digitised collection are the monitoring tools. In cultural aggregators, they are analytical tools developed ad hoc to allow the GLAM to monitor the status of metadata and the access and use of digital surrogates overtime.
Among the tools offered to GLAMs sharing their collection on cultural content aggregators, there are Google Arts and Culture dashboard and Wikimedia GLAM tools. The Google Arts and Culture dashboard allows GLAMs to customise their homepages and to analyse statistics on the views of the digitised cultural objects over the aggregator.
The Wikimedia GLAM tools are tools developed by the Wikipedia community in order to foster the GLAM-Wiki collaboration. Two of the most used tools are BaGLAMa 2 and GLAMorous, developed by Magnus Manske. BaGLAMa 2 shows the views of pages in Wikipedia and other Wikimedia sites containing digitised cultural objects. GLAMorous shows the number of Commons images used in other Wikimedia projects.

Notes

1. MoMA collection data on GitHub: https://github.com/MuseumofModernArt/collection.
2. Link to the "Directive on the re-use of public sector information": https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information.
3. ENUMERATE project on WIkipedia: https://en.wikipedia.org/wiki/Enumerate_(project)
4. The term "museum" has Greek origins. In Latin, the term "musaeum" means "institution of the muses". In ancient times, museum collected music, poetry and texts. Only in the th century, the term acquired its current meaning.
5. Other relevant institutions supporting the digitisation and the opening of cultural collections are the Joint Information Systems Committee (UK), the Digital Cultural Content Initiative (Canada), the National Endowment for the Humanities and the National Science Foundation (USA), Bill & Melinda Gates Foundation, the Ford Foundation and the William and Flora Hewlett Foundation.
6. OpenGLAM website: https://openglam.org.
7. A punched card is a piece of paper used to contain digital data represented by the presence or absence of holes in a grid.
8. Survey Report on Digitisation in European Cultural Heritage Institutions, (https://www.egmus.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-.pdf).
9. Link to the World Intellectual Property Organization (WIPO): https://www.wipo.int.
10. Link to the full text of the Berne Convention for the Protection of Literary and Artistic Works: https://wipolex.wipo.int/en/text/.
11. Inalienable rights refer to rights that cannot be given away or sold.
12. Open definition website: http://opendefinition.org/.
13. Creative Commons website: https://creativecommons.org.
14. Public domain license is equivalent to CC license. In most countries copyright ends years after the death of the author/s.
15. The Public Domain Review website: https://publicdomainreview.org.
16. Succession website: http://mtchl.net/succession.
17. GIF IT UP website: https://gifitup.net.
18. Trove website: https://trove.nla.gov.au/.
19. DigitalNZ website: https://digitalnz.org/.
20. The key figures for the museum collection are defined in the "Museum professionals in the digital era" report. It is available at the following link: http://www.project-musa.eu/wp-content/uploads///MuSA-Museum-professionals-in-the-digital-era-full-version.pdf.
21. Digital obsolescence is a condition where a digital resource is not readable because of its old format.
22. The Rijksmuseum website: https://www.rijksmuseum.nl/en.
23. The British Library website: https://www.bl.uk/.
24. The Deutsche Nationalbibliothek website: https://dnb.de/DE/Home/home_node.html.
25. Google Arts & Culture website: https://artsandculture.google.com/.
26. Microsoft In Culture website: https://www.microsoft.com/inculture/.
27. Hathi Trust website: https://www.hathitrust.org/.
28. Free Knowledge Foundation website: http://freeknowledge.eu/.
29. Open Knowledge Foundation website: https://okfn.org/.
30. Wikimedia Foundation website: https://wikimediafoundation.org/.
31. Europeana Foundation website: https://pro.europeana.eu/.
32. A cultural institution is subject to the laws of the nation where it is located.
33. What's the Picture website: http://www.whatsthatpicture.com/flickr/commons/.
34. Flickr The Commons website: https://www.flickr.com/commons.
35. WorldCat website: https://www.worldcat.org/.
36. Diplomatics is a scholarly discipline focusing on the critical analysis of documents, as well as their conventions, protocols and formulae.
37. Information science is a discipline that studies the collection, classification, manipulation, storage and retrieval of information.
38. In Lev Manovich founded the Cultural Analytics Lab.
39. Typically, cultural objects are scanned through flatbed, book or wand type scanners.
40. The Swedish library classification system (or SAB system) is used in Sweden to classify books and multimedia in the public sector.
41. Support Vector Machines (SVM) is a discriminative classifier used to categorize items starting from a labelled training dataset.

Mapping the Republic of Letters
by the Stanford Humanities Center, in collaboration with University of Oxford, Groupe D'Alembert (CNRS), CKCC and DensityDesign Research Lab, 2015. Website: http://republicofletters.stanford.edu.

"Mapping the Republic of Letters" is a research project which visually maps thousands of letters exchanged as part of the 17th and 18th century Republic of Letters. It investigates the correspondence of 13 influential intellectuals and their relationships with other relevant personalities of that time. Researchers developed static and interactive visualisation tools to display the social networks created by scientific academies and the physical networks created by travels. The project aims to create an open repository for metadata on early-modern scholarship, as well as provide guidelines for future data gathering.

2

Methodology

The thesis aims to define a set of guidelines for the design of cultural content aggregators that foster the access and usage of digital surrogates. The thesis follows a research through design. It consists of the following actions (Frayling, 1993):

- research of relevant case studies;
- development of a project and communication of the results;
- action research, a disciplined process of inquiry conducted for those taking the action;

The research through design approach facilitates the comparison between theoretical implications and real-world opportunities and constraints. It, in fact, is able to connect theory and experience and it has the dual role of validating the theory and giving a scientific structure to the experience (Frayling, 1993).
I developed two projects: Map the GLAM and GLAM Culture Hub. They allowed me to identify design solutions to facilitate access and use of collections within cultural content aggregators. The projects also allowed the cultural heritage stakeholders to engage in interactions with the collections that were not possible before.
Due to the large amount of data, some processes have been automated through the development of scripts - small programs created ad hoc for data acquisition and processing. However, most of the research adopts qualitative analysis.

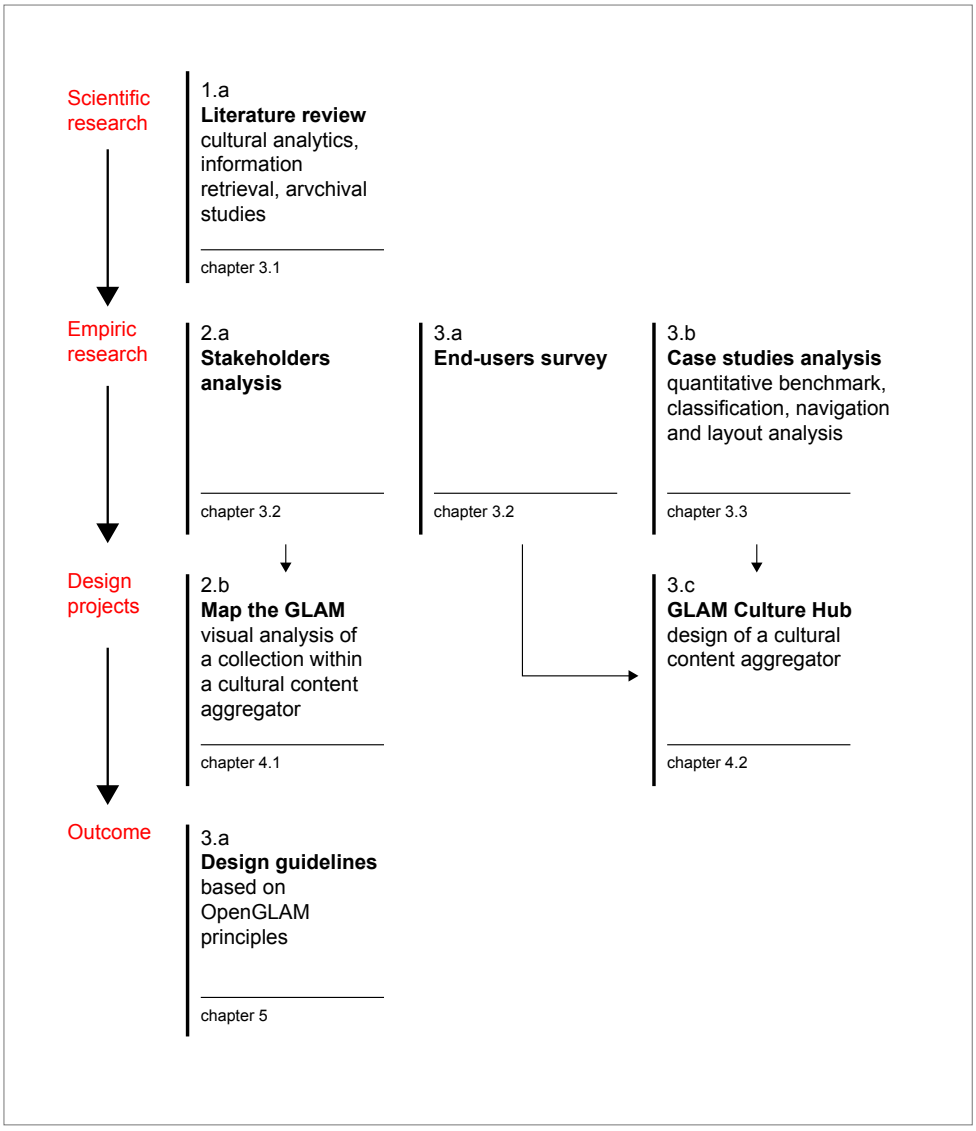


Fig. 2.1
Methodological path.

I carried on my research by applying the following methods (see fig. 2.1).

1. Evaluation
   a. Literature review.
2. Analysis
   a. Stakeholders' analysis.
   It includes people working for cultural institutions, such as librarians and digital curators, Wikipedia community and people working for Wikimedia Foundation.
   b. Map the GLAM project: analysis of a collection within a cultural content aggregator. I analysed the ETH-Library collection on Wikimedia Commons.
3. User research
   a. End-user survey.
   It aims to investigate end-users' motivations and methods of accessing the cultural content.
   b. Case studies analysis.
   I investigated five fundamental aspects of European cultural content aggregators, including the classification system and layout.
   c. GLAM Culture Hub project: design of a cultural content aggregator. It is based on the design of a high-quality prototype of a cultural content aggregator, with the aim to validate interface features that may foster the access and use of digitised collections.
4. Synthesis
   a. Definition of guidelines for the design of cultural content aggregators.

2.1 Literature review

I conducted a systematic literature review in order to understand the current research on subjects related to cultural content aggregators and digital archives. The literature investigates two domains: theInformation Science and the Digital Humanities. For the development of the design project, I also reviewed papers related to the use of cultural contents (in particular regarding Wikipedia) and the building of online communities.

Within the Information Science, I reviewed several publications concerning the fields of archival science, Information-Seeking Behaviours (ISB) and Information Retrieval

(IR) (see fig. 2.2). Archival science studies the ways human knowledge is organised and it is relevant because it provides an analysis of how people acquire information[1]. Archival science has given me an overview about past and present best practices for the organisation of a cultural collection.
Studies on Information-seeking behaviours focus on the strategies used by people to access information. They have been the most active research field in the area of Human Information Interaction (HII). Studies on Information-seeking behaviours allowed me to better understand the mechanisms an individual may use to access an online collection.
Finally, information retrieval studies refer to the study of different strategies (or models) to query an information system and receive results in the form of rank. They gave me the chance to know how computer-based information systems store and retrieve contents.

Within Digital Humanities, I reviewed publications concerning Cultural Analytics and the design of digital archives. Cultural Analytics applied computational techniques, such as statistical data analysis, data mining and information visualisation, to analyse cultural data. The datasets of Cultural Analytics essentially focus on cultural collections and user-generated multimedia files coming from the Web. Cul-

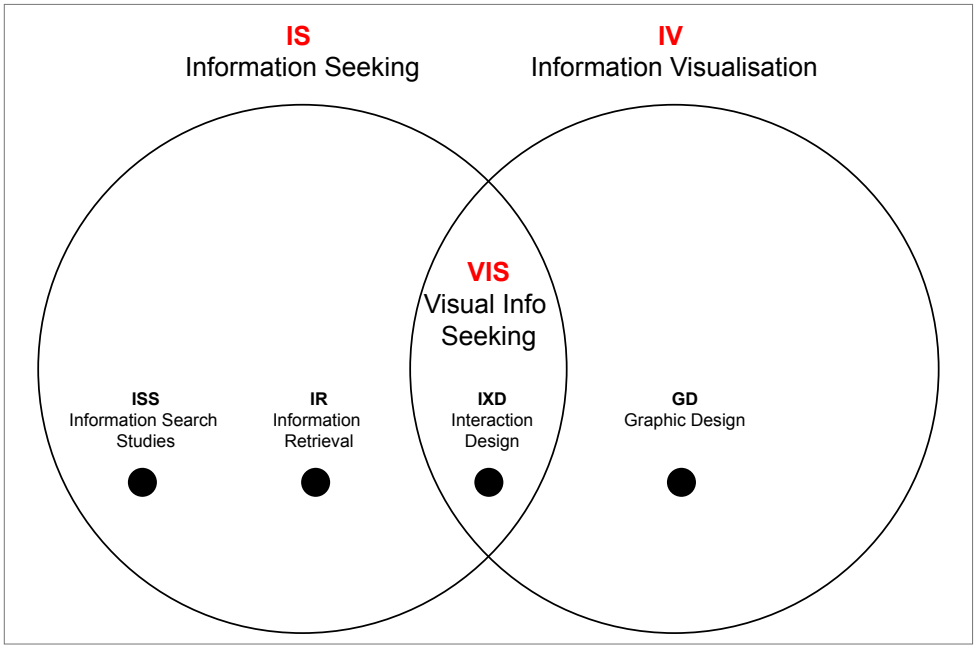


Fig. 2.2
Research context.

tural Analytics provided me with a set of aspects that could be investigated through the use of automatic data gathering and mapping processes.
Design research on digital archives is a field that encompasses competences coming from interaction design and data visualisation: these studies allowed me to gather design strategies and guidelines on how to visualise a cultural collection.

2.2 Analysis

The analysis consists of stakeholders' analysis, case studies analysis and an analysis of a collection within a cultural aggregator. The analysis aims to map the stakeholders of the open cultural content ecosystem and identify user interface requirements for the use of digitised collections within cultural content aggregators. During this phase I conducted a set of interviews to people daily using cultural content aggregators and I developed a set of data visualisations to analyse the status and the spread of a digitised collection within a cultural content aggregator. I also concentrate on some relevant digital archives emerging from the literature review. Among them, there are the Deutsche Digitale Bibliothek Visualized (DDB)[2] and the Deutscher Nationalbibiothek Visualised (DNBV)[3], by Urban Complexity Lab and Australian Prints + Printmaking[4], by Mitchell Whitelaw.

2.2.1 Stakeholders interviews

The stakeholder's analysis consists of a set of interviews to people working in the field of cultural heritage and that daily use cultural content aggregators. It aims to gather general information on how the cultural heritage system works and the interface requirements for every actor. I defined the different stakeholders through the analysis of the context of the open collections and the experience I acquired during the collaboration in research projects related to my thesis (see "Related projects" in the appendix). The stakeholder analysis includes only those individuals or groups who daily use cultural content aggregators. Therefore, policy makers, legislators and other people dealing with digital collections in general are excluded. I interviewed nine people belonging to three types of stakeholders: GLAMs, nonprofit organisations and communities.

- GLAMs
  - Michael Gasser, Head archives of ETH-Library
- Nonprofit organisations
  - Florence Devouard, Chair of the Wikimedia Foundation (2006-2008)
  - Alex Stinson, GLAM-Wiki strategist, Wikimedia Foundation
  - Barbara Fischer, Curator for cultural partnership, Wikimedia Germany
  - Dario Crespi, Coordinator of the Lombardy region, Wikimedia Italy
  - Jenny Ebermann, Executive director, Wikimedia Switzerland
  - Ilario Valdelli, Community manager, Wikimedia Switzerland
- Communities
  - Erica Litrenta, Community liaison, Wikimedia Foundation
  - Stefano Dal Bo, Ysogo (nickname), Wikipedian

With every stakeholder we essentially discussed about three main aspects: the role it has within the open cultural content ecosystem, issues and potentialities related to its work within cultural content aggregators and its nice to have. The information acquired has been reported in specific paragraphs within chapter 3.2 according to the following order.

- Stakeholder type
- Role and function within the cultural content ecosystem
- Goals
- Knowledge/expertise
- Relationships with other stakeholders
- Problems and user interface requirements

2.2.2 Analysis of a digitised collection

I analysed a cultural collection released on a cultural content aggregator to understand the dynamics that drive the access and the use of digital surrogates. I have selected Wikimedia Commons - the media repository of Wikipedia - as cultural content aggregator because it has millions of users and publishes in open source data about the collection usage. I have also chosen the collection of ETH-Library in Zurich, because it is one of the biggest available on Wikimedia Commons. Through the use of data visualis-

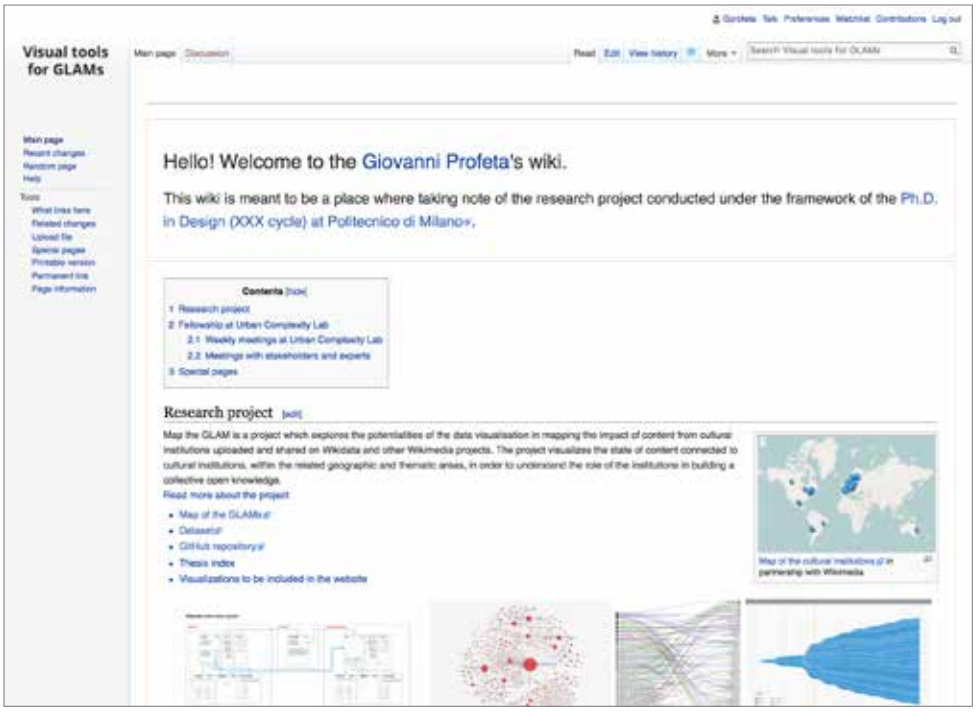


Fig. 2.3
Wiki of the analysis of the ETH-Library collection. It was online during the analysis phase.

ation, I tried to map the status and the spread of ETH-Library collection over Wikipedia and its sister projects. Due to the huge amount of data, I had to code several programs in order to gather and elaborate them quickly. I have also defined a set of visual protocols for data collection and display to facilitate the replication of the analysis, by adopting the following strategies in order to optimize the time at my disposal.

- Plan works and deliverables;
- Clearly define the research goal to skip useless information;
- Automate data gathering and quantitative analysis;
- Elaborate quick visual explorations of data;
- Alternate more convergent than divergent design phases;

I have implemented my personal wiki to document the whole analysis phase, and to better understand the technicalities a wiki software adopts (see fig. 2.3). The source code of the scripts I implemented is released under open license on GitHub. I have also shared the methods I adopted in the project on Wikimedia Meta. After sharing it, I have received

feedback from the Wikipedia community and some tips to correct minor errors. Feedback on visual analysis was useful to define some interface requirements.

2.3 User research

The user research consists of an end-user survey, a case studies analysis and the design of a cultural content aggregator. The analysis aims to identify interface requirements and user motivations for using cultural digitised collections within cultural content aggregators. During this phase, in particular during the case studies analysis, I adopted some methods to map the informative and visual structure of existing cultural content aggregators coming from Europe. This allowed me to gain some expertise for the design of a cultural content aggregator.

2.3.1 End-user survey

I conducted the end-user survey to investigate reasons and methods of accessing content among people who often use cultural content aggregators. The survey is an ethnographic study consisting of a list of questions about the experience of using these kind of web platforms. I sent the survey via email to a set of people who published online some images from the ETH-Library collection. I gathered this set of people starting from the 100 most used images on Wikimedia projects (obtained through the use of scripts developed in the framework of the Map the GLAM project) and searching for them on the web via TinEye - an image search engine[5]. The survey consisting of a list of the following questions:

- How did you access the picture?
- It was easy to find the image you were looking for?
- What kind of issues do you usually experience when you look for an image in an online digital archive?

This method allowed me to gather information about the use of cultural content aggregators and to collect interface requirements according to the different typologies of end-users, such as writers and photographers. Since this method is based on the republish of digital surrogates online, it was not possible to reach end-users such as scholars and students. However, I collected information about their needs through the stakeholders analysis.

2.3.2 Case studies analysis

The case studies analysis aims to identify issues and useful features of existing cultural content aggregators. It is based on a selection of cultural content aggregators contributing to Europeana[6] - a cultural aggregator funded by the European Community. Every case study is selected according to the following requirements: collecting digital images[7], providing users with a graphical user interface (API-based cultural aggregators are excluded) and using the English language. According to these requirements, 14 case studies, including Europeana, have been selected. The goal of the case studies analysis is to identify the fundamental interface features of cultural content aggregators.
I investigated five fundamental aspects of the selected cultural content aggregators: content, classification system, information architecture, access modes and layout.

Content analysis
In the content analysis, I examined the metadata belonging to a cultural item. Since the metadata concerning the images are essential to organise them within an information system, this analysis has been useful for the development of the following classification system analysis.
Different aggregators may use different terms for the same metadata. In order to o overcome any misunderstandings about their meanings, I used the following univocal vocabulary:

- Title;
- Date (or period);
- Author;
- Description;
- The owner (the cultural institution);
- License of use;
- Category; it refers to one or more terms that frame the artwork within a specific thematic area, the level of specificity of the categories usually can be mapped through a tree;
- Tag (or keyword); it refers to terms that describe the artwork, they can be multiple and have different levels of specificity;
- Subject (the object of the representation);
- Typology (or material), the medium of the cultural object;
- Technique, the artistic technique used to create a cul-

tural object;
- Colours (or colour palette);
- People involved, who have collaborated in the creation, preservation or restoration of the work;
- Size, dimensions of the cultural object and the related digital item;
- Location (or provenance), the physical place where the cultural object is found or stored;
- Sector (or collocation), the location of the cultural object within the collection, which may consist of an alphanumeric code;
- Identifier, which refers to a univocal alphanumeric code that identifies the cultural object;
- Exif data[8];
- Others metadata, such as the depicted place, the people involved and the file format.

Classification system analysis
In the classification system analysis, I examined how cultural content aggregators classify items. This is a qualitative analysis based on the observation of the tools to access the collection. Since the classification of content is essential in basis every information system, I tried to understand how cultural content aggregators organize classes of images and how these are connected. Even though I know the way technical databases store information, I did not consider system backend, but what emerges from the interface.

Information architecture analysis
In the information architecture analysis, I examined the links between the pages of the content aggregators. Links to previous pages that can only be done using the back button of the browser have not been taken into consideration. The information architecture analysis is based on a navigation analysis method coming from the literature review (Kreiseler et al., 2017). The goal of the information architecture and the following analysis is to identify the strengths and weaknesses of the existing user interfaces.

Access modes analysis
In the access modes analysis, I considered, in qualitative terms, the interaction patterns to access digital surrogates with reference to the use of visual displays both to facilitate the analysis of the entire collection and to access specific

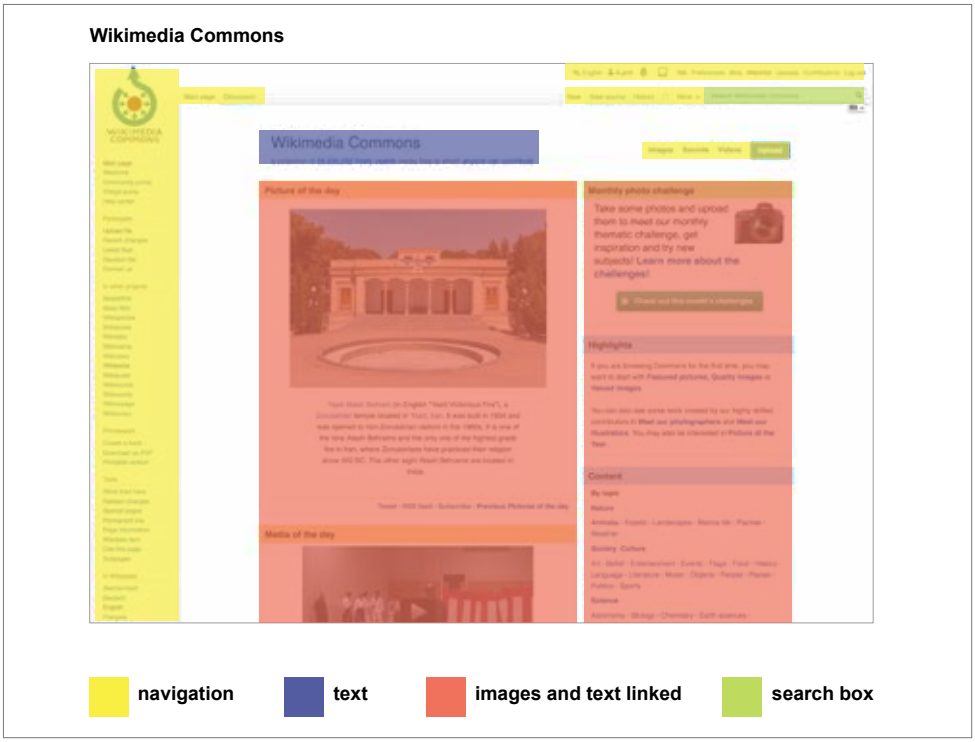


Fig. 2.4
Layout analysis of Wikimedia Commons.

digital surrogates. In addition to the well-known search and browse access modes, I also considered the explore mode. The explore mode is also defined by the existing research as "visual information seeking (Shneiderman, 1996) and "engage mode" (Stiller, 2012).

Layout analysis
In the layout analysis, I examined, in qualitative and quantitative terms, the distribution of contents over the aggregator pages (see fig. 2.4). The analysis is based on a wireframe analysis method coming from the literature review (Kreiseler et al., 2017). I used Full Page Screen Capture - a Google Chrome plugin - to extract screenshots of the cultural aggregators, together with a vector design software to redesign the wireframes.

2.3.3 Design of a cultural content aggregator

This last research method consists of an evaluation by end-users of a cultural content aggregator interface through a high-quality prototype I realised. I have designed a cultur-

al content aggregator to validate interface features that may foster the access and the usage of digitised collections. The design of the prototype complies with requirements based on the literature review, the analysis of the aggregators, the analysis of the ETH-Library collection and the attitudinal survey. It displays contents coming both from Wikimedia Commons and Europeana.
The survey is based on the literature about user centred design (IDEO, 2015), web usability (Krug, 2000) and the System Usability Scale (SUS) - a set of questions for measuring the usability of an interactive system. The survey includes open and closed questions concerning visual and interface elements.
The survey was launched in two rounds. In the first week, I re-contacted people who had replied to the attitudinal survey. I also contacted via email several people working or collaborating in the field of cultural content, including the Wikipedia community.
In the second week, I published the survey on my personal profiles of Twitter and Linkedin and I wrote an article for the newsletter of the Wikipedia community.
At the end of the first week, I slightly revised the introduction of the survey to make it easier to understand for users. I also made an explanatory video of the main aggregator features.
The survey gathered both qualitative and quantitative data. People who replied to the survey are professionals working in the cultural field, the Wikipedia community, designers and photographers.

2.4 Synthesis

I defined a set of guidelines for the design of cultural content aggregators, which are the final outcome of the research. They try to synthesise in five chapters the knowledge acquired through scientific research (the literature review) and empirical research (the analysis of case studies and the user research). The goal of the design guidelines is to provide designers with guides on how to encourage the access and use of cultural content within cultural content aggregators.
The design guidelines tackle four main aspects: classification, access, navigation and use of digital surrogates. The following list briefly shows all the chapters in which these aspects are discussed.

Classification
- 3.3.1: analysis of classification systems of existing cultural content aggregators;
- 4.3.1 – 4.3.2: two design projects;
- 5.1: design guidelines;

Access
- 3.3.2: analysis of access tools of existing cultural content aggregators;
- 4.3.2: GLAM Culture Hub project;
- 5.2: design guidelines;

Navigation
- 3.3.3: analysis of navigation tools of existing cultural content aggregators;
- 4.3.2: GLAM Culture Hub project;
- 5.3: design guidelines;

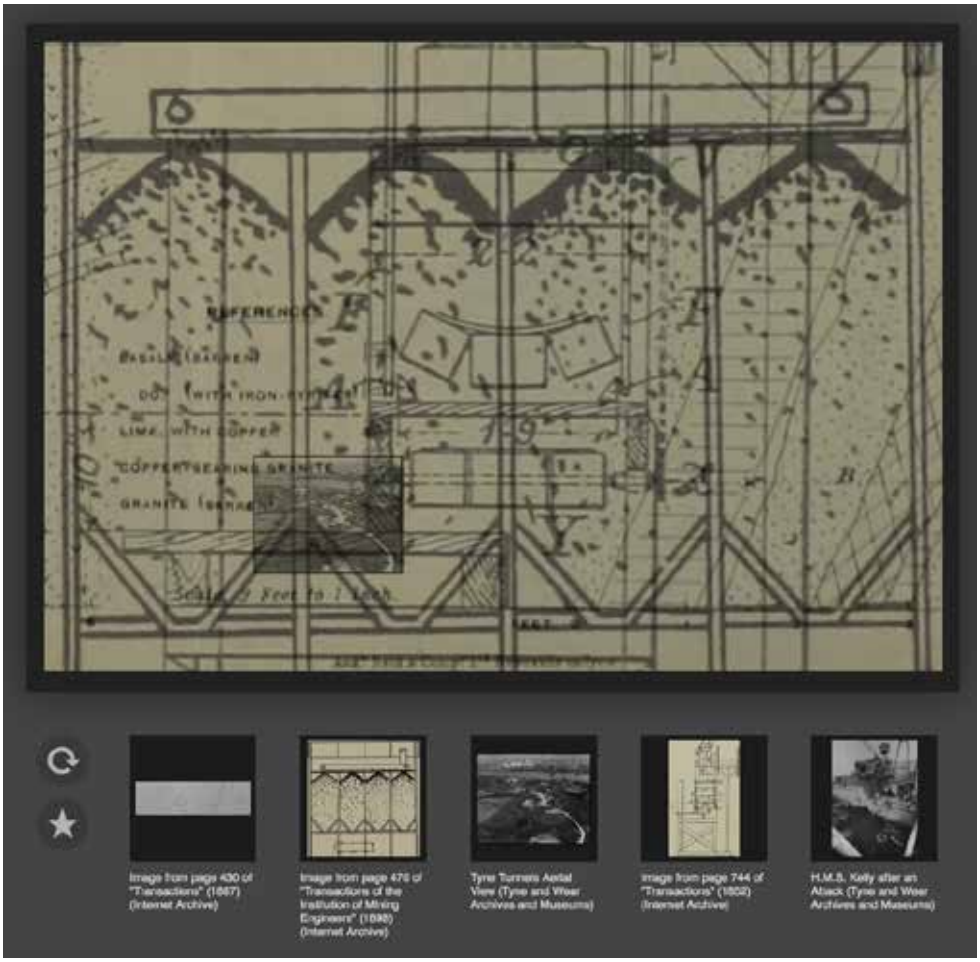
Use
- 3.3.4: analysis of tools for using digitised collections within existing cultural content aggregators;
- 4.3.1: Map the GLAM project;
- 5.4 – 5.5: design guidelines;

The design guidelines are based on OpenGLAM principles - a set of five guidelines developed by the Open Knowledge Foundation, which provides cultural institutions with strategies aimed to open their collections.
I transposed the principles concerning the cultural institutions into design guidelines, according to a complementary principle. OpenGLAM principles give GLAMs the content requirements and my OpenGLAM Design Guidelines provide designers with interface requirements. For conceptual reasons, I transposed the principles number 3 and 4 of OpenGLAM into the points 4 and 3 of the design guidelines (see chapter 5).
Every guideline presents contents according to the following structure:

- Introduction to the topic;
- Context and user requirements;
- Design strategies at a high level;
- User interfaces elements.

Notes

1. I have chosen to use the term “acquiring information” instead of the most used “seeking information”. This is due to the fact that recent studies have shown that most of the information acquired daily is obtained without a conscious attempt to search for it.
2. Deutsche Digitale Bibliothek Visualized website: https://uclab.fh-potsdam.de/ddb.
3. Deutscher Nationalbibiothek Visualised website: https://dnbvis.fh-potsdam.de/
4. Australian Prints + Printmaking website: http://printsandprintmaking.gov.au.
5. I used TinEye because it allows me to search for images programmatically.
6. Europeana website: https://europeana.eu/portal/en.
7. Even though cultural content aggregators collect several digital surrogates, including videos and 3D models, I focused on images because they are currently the more widely diffused surrogates.
8. Exif data are metadata produced by the process of digitisation of a cultural object, for example the type of image compression, file size (usually in kb or mb), the type of camera used and resolution (in dpi).



Succession - Digital fossils from an industrial city, by Mitchell Whitelaw, 2014. Website: http://mtchl.net/succession/.

"Succession - Digital fossils from an industrial city" generates digital fossils by compositing images documenting the impact of the industrial revolution to the city of Newcastle. Each image is generated by randomly selecting five items from a set of around 2000 coming. The source images come from the Flickr Commons, largely Tyne & Wear Archives and Museums, the Internet Archive and the British Library, and are released under a Creative Commons license. The project aims to foster reflection on the past and the future about the industrial culture.

3

User interfaces for open digitised collections

Cultural institutions release collections under open licenses through multiple channels: official websites, digital archives, content aggregators and via API (to allow the creation of web and mobile applications). In recent years, cultural content aggregators have become a powerful resource for scholars, writers and students. By aggregating open collections coming from several cultural institutions, content aggregators offer the opportunity to quickly access, analyse and reuse a vast cultural heritage. Content aggregators also provide end-users with tools for contributing information, organising content and generating new knowledge.

At the core of cultural content aggregators (as well as digital archives), there is a shift from focusing on static, physical objects towards dynamic, digital surrogates. A shift equally from seeing cultural objects create stable hierarchical organisations to situating digital surrogates within horizontal fluid networks. For cultural institutions, this shift requires moving away from identifying themselves as passive guardians of cultural heritage to actively shaping collective memory (Cook, 2001). For designers, it opens novel design challenges concerning the access to heterogeneous and visually rich collections (Chen et al., 2014). Designers must deal with classification and technological aspects in a multidisciplinary context. Due to their expertise, they may have a primary role in creating cultural content aggregators, not only with regards to the design of the information system but also for the contribution in making sense and generating new insights from the collections.

Currently, public and private institutes show a growing interest in applying new technologies to vast cultural collections. Among them, there are Artificial Intelligence (AI), Machine Learning (ML) and Augmented Reality (AR).
An interesting project using machine learning technologies is "LIFE Tags"[1] (see fig. 3.1). Developed by Google, "LIFE Tags" is an interactive encyclopaedia featuring millions of images from LIFE magazine[2]. The project automatically classifies all the images published weekly between 1936 and 1972 and monthly from 1978 to 2000 into a catalogue based on thousands of labels[3]. Users can easily navigate the digital archive by browsing image categories and labels. A new random layout is generated at each visit, with different highlights and titles and every label is linked with its related Wikipedia article to provide users with additional information.

Although technologies are becoming more and more performative by allowing quick access to millions of items, several studies demonstrate that only a small part of online cultural content is used by end-users (Borowiecki and Navarrete, 2016). This is due to reasons related to the nature of the channel (the Internet), the content and the user interfaces.

The channel is decisive in defining the access modalities to the content. Generally, the usage pattern of digitised contents, such as images and books, presents a long tail[4] where

Fig. 3.1
Life tags, Google, 2018.

few items are the most popular and the majority of the content remains obscure. The Internet has led to a lengthening and thickening of the long tail (Anderson, 2006). Thus, popular items drive the attention of end-users towards certain contents in spite of others. In the end-user selection and eventual use of cultural content, both continuous technological improvements and update of contextual information (also referred to as "information curation") have a primary role (Ginsburg and van Ours, 2003). Technology provides users with tools to retrieve plenty of content. Its goal is to make every single item accessible and available without any restriction or bias. Contextual information refers to the information which provides a temporal and semantic framework to a cultural object. This information is given at a higher level than metadata, and both cultural institutions and end-users can generate it. It was demonstrated that web communities are influential in the distribution of access and usage of digital surrogates (Borowiecki and Navarrete, 2016) (see fig. 3.2).

The content consists of the digital surrogate and its metadata. The quality of the article influences the number of views positively (Navarrete and Borowiecki, 2016). However, many content aggregators present contents consisting of low-quality images and few or inaccurate metadata.

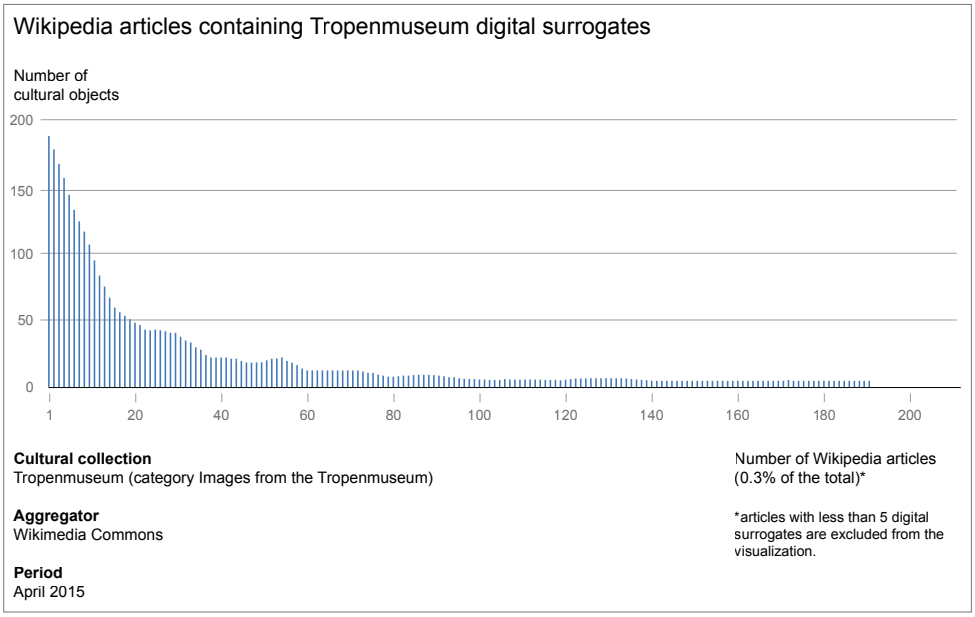


Fig. 3.2
Pattern of usage of the Tropenmuseum collection on Wikipedia. Redesign of the chart based on the paper by Navarrete and Borowiecki, 2016.

The analysis of contents provided by GLAMs to Wikimedia Commons showed that part of them does not respect the aggregator guidelines regarding the educational purpose[5]. Furthermore, due to the data entry made by many people, image classification and metadata are often inconsistent.

The design of user interfaces impacts the user's experience. Some studies claim that many museums do not focus enough on users and their related needs (Farber and Radensky, 2008; Roberto, 2008; Pruulmann-Vengerfeldt and Aljas, 2009). Designing without keeping the user in mind may lead to several usability issues, including the ones related to the access to digital surrogates. Current interfaces of digital are almost entirely search-based. Search boxes are familiar and powerful interface tools, but they are thought for a small audience that understands a collection's content and can query it effectively (Whitelaw, 2012).
In order to overcome usability issues and foster the access and usage of cultural contents, the design of content aggregators must consider the end-users' needs and skills. Furthermore, designers should change the concept behind content aggregators: from static depots of digitised content to dynamic tools aimed to support research and dissemination.

Digital archives and cultural content aggregators were born because of different needs. A digital archive is mainly conceived as a way to explore collections for scholar purposes, while a cultural content aggregator is mainly thought as a collector of multiple collections in order to allow their access and use for dissemination purposes.
However, it is possible to deepen the principles that govern the classification and management of cultural collections within the digital archives, to understand the methods of accessing and using cultural objects within cultural content aggregators.

3.1 Research in accessing collections and information

The study on the interfaces of cultural content aggregators should consider the research in the fields of information science and cultural analytics. Information science (IS) is associated with the use of computers and technology in general. It is a scientific field that studies the analysis, classification,

storage and retrieval of information. In the contest of this user-centred work, information science is relevant for the studies concerning the organisation of information and search modes. Cultural analytics, on the other hand, is relevant for the introduction in digital cultural archives of interactive visual models[6] as analysis and access tools.

3.1.1 Modes of information acquisition

The concept of information need refers to the set of information which is necessary for a user to complete a task, objectively. How the user asks for this set of information is called Information demand (Fidel, 2012). An information system must handle this request by ensuring a good match between information need and information demand.
The information allows to fulfill several tasks, such as getting insights on a past event and understanding the current situation. However, people need information for six primary purposes, each of them requiring different information solutions:

1. Getting answers to specific questions in the form of who, why, what, where, when and how (fact-finding function);
2. Keeping up to date (current awareness function);
3. Investigating a new field in-depth (research function);
4. Achieving a background understanding of a topic (briefing function);
5. Obtaining novel ideas or stimuli (stimulus function);
6. Looking for tidbits of information for fun (recreational browsing function) (Nicholas, 2009).

In the context of cultural aggregators, information about digital surrogates is indispensable because they allow its access and evaluation.
For several years, studies conducted in the field of Information Search Behaviours (ISB) considered the process of gathering information to fulfill a task as a conscious, human activity. Recently, researches show that seeking information can also be done unconsciously or automatically, due to the fact that humans have searched for information over millions of years of evolutionary development. Thus, they have carried out the information search in a completely natural and unconscious way by using passive and sampling behaviours (Bates, 2002).

Even though people acquire information not only through a search, the term "information seeking" is still widely used in scientific publications. However, it also refers to unintentional ways of getting information.
There are several modes of information acquiring, defined according to the degree of cognitive effort and focus. The cognitive effort can be "active" or "passive", whether the individual does anything actively to acquire information, or he is passively available to absorb it (see fig. 3.3). The focus can be "directed" or "undirected", whether the individual seeks particular information, or he is randomly exposing himself to information. Starting from this schematisation, four modes of information acquiring emerge: searching, browsing, monitoring and being aware. Monitoring and searching are the ways people find information that needs to know, whilerowsing and serendipity are the ways people find information they do not know they need to know (Bates, 2002).

Searching
Searching for information is an active process that involves the interaction between the user and the information (Belkin et al., 1993). When an individual seeks information, he purposely looks for information to answer questions (Bates,

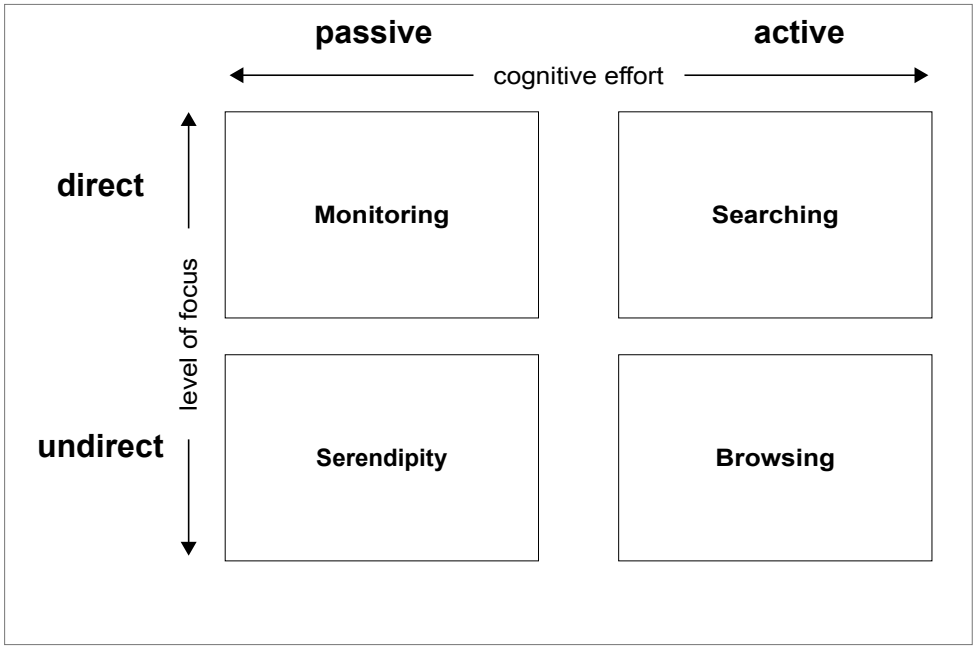


Fig. 3.3
Modes of information acquisition.

2002), to support decision-making (Fidel, 2012) and to fulfill an information need.
The searching mode includes two main types of search strategies: active and passive strategies. Active strategies entail a direct interaction between the seeker and the target of the search by enabling different tactics, such as asking someone for information. Passive strategies involve acquiring information through unobtrusive observation, such as being in CC on an email or eavesdropping on a conversation (Ramirez et al., 2002).
Due to a large number of search strategies, scholars have proposed several distinctions. The most relevant are the distinctions between seeking and searching and between analytical and empirical search. Seeking concerns the variety of methods people employ to discover and access to information resources, while searching is a subset of information-seeking regarding the interactions between a user and computer-based information systems (Wilson, 1999).
In an analytical search, an individual explores the information needs on one hand, and information system capabilities on the other, in order to translate the need into a query in the system's language. In an empirical search, this is a shortcut to the analytical one, an individual uses rules and tactics that were successful in the past (Fidel, 2012).
An excellent interface for searching for information provides search options to narrow the information and a set of relevant results for the user query.
Nowadays, due to the dramatic increase in data and content humans generate, searching is the dominant way of acquiring information. The ability to search for information properly is recognised as a must-to-have soft skill in several public and private sectors.
Information seeking strategies comprise interactive intentions and retrieval tactics. Interactive intentions refer to sub-goals that a user have to achieve in the process of accomplishing his search task. Retrieval tactics refer to techniques the user applies to interact with data and information (Xie, 2007).

Browsing
Browsing refers to the act of actively expose ourselves to possibly new information with or without a specific need. People can browse both physical and digital spaces, such as libraries and websites. While browsing, they are engaged in a series of glimpses (Bates, 2007) and scan items of potential

interest (Rice et al., 2002). This exposition may or may not lead the individual to (physically or conceptually) acquire the item (Bates, 2002). The final goal of browsing concerns mainly the pleasure and not the task (Toms, 2000).
A good interface for browsing information provides browsable, serendipitous and enjoyable information spaces (Rimmer et al., 2008). It is full of potential objects of interest that the eye can take in at once, also called "massively parallel glimpse" (Bates, 2007).
This browsing strategy has been identified several years before computers began to be used for information retrieval (Fidel, 2012). Currently, it is the most pervasive, used strategy in information acquisition. Even though it is a strategy on its own, it can also be adopted while other modes are employed.

Monitoring
Monitoring is the activity of surfing a source of information to see what it contains, without having an information need. As well as in the browsing strategy, people can monitor information both in physical and digital spaces. Even though this mode brings unpredicted information and serves a not ongoing decision-making process, it is not a completely random act. An individual chooses to focus his attention towards specific elements (Fidel, 2012) by maintaining back-of-the-mind alertness for items that can be of interest and for answers to questions they might have (Bates, 2007).
Interfaces that support information monitoring provide an extensive display of what is available to allow individual selecting possible items of interest.
Despite its popularity, monitoring is not officially recognised as a research topic. However, market researchers study it to help businesses to advertise their products, so potential consumers can encounter them in a pseudo-casual way.

Serendipity
Serendipity is the act to "bump into" information when people are not seeking at it, but it might solve a problem in the future. Casually finding a telephone number or useful information when reading for pleasure are examples of serendipity (Fidel, 2012).
This acquisition mode does not rely on specific interface characteristics. The acquisition process is unpredictable; thus, studies on serendipity mostly focus on individuals report on previous experiences.

Serendipity has several synonyms, such as casual information-gathering, passive information-seeking, incidental information acquisition and the accidental discovery of information. The research established that people casually acquire information more frequently than has been assumed. Serendipity gives people around 80 per cent of all it knows (Bates, 2002). Some of them primarily rely on this mode (Fidel, 2012).
People acquires information daily, through active and passive access modes. The choice of an active acquisition mode (searching or browsing) depends on four conditions:

1. The time an individual has to acquire information;
2. The prior knowledge required for the employment of a mode;
3. The cognitive effort required when using a mode;
4. The human remembering required for the use of a mode (Fidel, 2012).

The searching mode requires some knowledge to be productive. In particular, an individual has to have some knowledge of the search system and the subject domain. In comparison with the browsing mode search time is limited and thinking, analysis and some search in memory are required. The search mode is more useful when an individual is searching for something new, such as a new topic or a variation of a topic in a new information system (Fidel, 2012).
The browsing mode is the most comfortable way to apply, because it does not require previous knowledge, nor much thinking or remembering. However, it is less efficient that the searching mode because it rarely leads to the desired information directly, and therefore, it requires more time. The browsing mode is more useful when an individual has little knowledge of the subject domain but has no strict time limits (Fidel, 2012).

As discussed before, people usually interact with information in order to use them when they need to make decisions or solve problems. Therefore, using information is a fundamental aspect of information science and its related disciplines. The most prolific area of use-related research is bibliometrics, which is based on the analysis of citations in scholarly publications. When one document cites another, a set of relationships between the two can be assumed: this may indicate that both documents have the same topic and,

in combination with other documents, they may reveal several types of associations among the authors.
Using information is a highly complex task because it brings constraints (such as limited time and restricted access modalities) and demanding cognitive process (such as acquiring, filtering and evaluating information) (Fidel, 2012).

3.1.2 Visualisation as an access tool

Information, by its definition, is communicated. Thus, when information is represented, it takes a particular form (Fidel, 20012). Among the different forms of representation, there is information visualisation (Tufte, 2001). Recently, visualisation is acquiring a primary role as a tool to access information in several academic and non-academic fields, including the cultural sector. Data visualisation covers both the scientific and the design domain. It is the graphical representation of information and data by using visual elements, such as charts, networks, and maps. It is demonstrated that visualisation can integrate humanities research processes while supporting casual explorations by general-interest readers (Hinrichs et al., 2015).

Visual displays of information are cognitive artefacts that complement and strengthen our mental abilities. Among these, visual displays can increase working memory, facilitate search, support perceptual inference and enhance detection and recognition of patterns (Meirelles, 2011; Card 1999). The computer has been responsible for massive advances in the field of data visualisation. It makes possible the storage of vast amounts of data, allows the interaction with data and provides high-resolution graphic displays (Spence, 2007, p. 6).
In recent years, there is a growing body of research and projects about the use of data visualisation as a navigation tool to explore cultural content (Manovich, 2007). This is mainly due to the need, by scholars and learners, to have artefacts able to provide a visual representation of properties of digital surrogates.
Visualising cultural content is a mapping process. Descriptive metadata (also called content-based metadata), such as style and genre and structural metadata (also called visual and semantic-based metadata), can be represented in 4 dimensions (X, Y, Z, time) (Ushizima et al., 2012). However, the metadata of the cultural object have more than four

dimensions. Thus, designers and cultural institutions have to choose which dimensions are to be used and which one have to be omitted. This choice leads to the rise of questions about mapping the selected dimensions: who has the power to decide what kind of mapping have to be used? What dimensions are selected? What kind of interface is provided for the user? (Manovich, 2002). Generally, designers carry on workshops of co-design with stakeholders to answer these questions. The outcomes of the design activity can be curated or procedural representations, or a combination of the two (Whitelaw, 2015). Procedural representations are visualisations generated through a formal (usually computational) process operating on the collection data. Curated representations are hand-crafted representations that reflect a curatorial intent. Current cultural collections mostly use procedural representations in combination with search, browse and facet access modes.

Recent research has sought to extend the search mode of information acquisition. It proposes the "exploratory search", arguing that collections interfaces should be more dynamic and interactive, by enabling users to explore through selecting facets (or browsing), rather than requiring a specified query (Marchionini, 2006). However, current digital collection interfaces use faceted search as exploratory access paradigm only after traditional query. Thus, search mode remains the primary way to access collections.
In digital archives, a search-centred user interface responds to the need of many users to access specific content (Fantoni et al., 2012). It also reflects a functionalist philosophy, which has been strongly criticised by some other studies. Some of them, coming from the information science domain, recognise the limits of explicit information and the inarticulate feeling of a query (Belkin et al., 1982). Some others, coming from the humanities, show that browse features are highly valued by non-expert visitors of online collections (Lopatovska et al., 2013). They also show that functional models of task-goal are inappropriate, because distraction, engagement, flow experience and pleasure-driven activity are not goal-oriented, but motivated by the process itself (Drucker, 2013). Information acquisition, through the means of visual representation, emerges as a complex, open-ended process, rather than a goal-oriented transaction (Marchionini, 2006). The task-based paradigm embedded in search mode does not satisfy the needs of the users accessing digital archives,

but other models are emerging: they are not alternatives but rather complementary to the search mode. Among them, the "Information Flaneur" model is the most relevant: it belongs to the serendipity mode and envisions a person who explores digital archives through several information modes, including exploratory browsing, leisurely curiosity and aesthetic pleasure. Information Flaneur comes from the French masculine noun "flaneur", which refers to a urban character who feels at home in the boulevards, arcades and cafes of Paris in the 1840s. The introduction of the Information Flaneur aims to promote a shift from negative concepts, such as the information need and the task-goal models, towards positive information experiences (Dörk et al., 2011).
The information flaneur model requires digital collection interfaces which provide rich overviews and foster serendipity. They are called "generous interfaces" and are in contrast with search-based interfaces, mainly designed for expert users and able to query it effectively. A generous interface matches the abundance of digital collections and the generous ethos of the institutions holding them. It visually represents the scale and richness of the collection, offers multiple ways in and enriches interpretation by revealing relationships among items. By using navigable representations, they invite exploration and support browsing (Whitelaw, 2015).
A key challenge in designing generous interfaces is compression, that is the need to represent extensive, diverse collections in a compact, browsable form. A set of principles for designing generous interfaces have been formulated in order to overcome this and other related challenges.

1. Show first, do not ask. Non-expert audience does not like search boxes. Generous interfaces should encourage easy enquiry.
2. Provide rich overviews of collections. Do not dictate user journeys either narrow the information experience.
3. Provide samples. Use samples, such as image thumbnails to represent and characterise a digital collection. Samples provide rich contextual cues and invite exploration.
4. Provide context by displaying the structure and relationships within the collection. Both item-to-collection relationships and item-to-item relationships are important.

5. Share high quality, primary content and provide access to high-quality primary content, good images and stables URLs (Whitelaw, 2012).

In a generous interface, interactivity plays a fundamental role. It is used to enable users to immerse themselves in the historic situation an object gained significance from (Stiller, 2014) and traverse the extent of the collection. A study on interactive visualisation proposes some general guidelines concerning the design of interactions.

1. Relate facets and items: the interface should use visual variables, such as colour, position and shape, to convey relationships among items and their shared facet values.
2. Encourage pivoting: every visual element should provide simple pivoting interactivity to move between sets of resources. Pivoting steps should be simple to trigger and reverse.
3. Make view changes gradual: transitions between interface states should explain how and why information changes. Though feedforward[7], the interface should indicate some of the changes.
4. Aim for inviting aesthetics: the overall appearance of the interface should follow a consistent, easy to understand and enjoyable visual language.
5. Situate in the Web context: the interface should be positioned in the context where information seeking already happens, for instance, the web browser. Common uses and features of the Web should be supported to benefit from the strolling experience (Dörk et al., 2012).

Changes in established conventions for digital cultural collection interfaces involve several challenges across multiple domains. Generous interfaces are more technically demanding in terms of data manipulation and computational load. Moreover, the need for overviews involves new data requirements so, collections' data storage must shift from search-oriented databases to a new database, which is more suitable for these interfaces (Whitelaw, 2012).

To sum up, studies in information science, cultural analytics and digital archive interfaces are complementary to each other. The first mostly focus on the search mode, while the second bases its theoretical assumptions and design pro-

jects on the serendipity mode.
This thesis agrees with the two positions by considering them as two different approaches aiming to solve the same design problem. The study tries to extend both the studies concerning information science and cultural analytics in order to create an integrated approach. It actively considers the primary audience of cultural content aggregators, including scholars and other expert users. It aims to offer design guidelines for the coexistence of multiple content access modes. The integrated approach emerges from three main findings of the literature review:

1. the research demonstrated that a single access mode is not sufficient for a variety of users of cultural content (Dörk et al., 2011);
2. empirical studies showed that dual modes for the information acquisition are more efficient and enhance the findability of the item's findability (Lin et al., 2016);
3. the use of visual displays makes easy to get an overview of the collection (Whitelaw, 2012), while the use of search boxes narrows the research to a few results (Fidel, 2012). Thus, by integrating the two interface elements, users can quickly choose to get information at the scale they prefer.

The proposal of an integrated approach also reflects the multidisciplinarity of figures required to design user interfaces for open digitised collections including domain experts, information designers and computer scientists.

3.2 Stakeholders and user requirements

Around digitised collections released under open licenses within cultural content aggregators there is the interest of multiple stakeholders. From the analysis the following stakeholders emerge: cultural institutions, nonprofit organisations, communities based on open content, writers and communication agencies. Among stakeholders there are also scholars and students[8]. Every stakeholder has his own expertise, goals, role within the open cultural content ecosystem and requirements[9]. The stakeholder's requirements may be related information, functional and non-functional needs. The designer works for nonprofit organisations and should consider the requirements

of every stakeholder, and in particular those regarding end-users (writers and communication agencies).

Cultural institutions

With the term "cultural institutions" we refer specifically to the responsible for the digital archives, or digital curator[10]. He has the fundamental tasks of ensuring the access and maintenance of digital surrogates. Digital curator also has a crucial role within the cultural open content ecosystem because it establishes the objects to be digitised, the digital surrogates to be released and the licenses to be adopted. Thus, this has a strong impact on the content that will be produced by end-users on the basis of digital surrogates. In addition to the digital curator there are other figures who deal with the digitised collections, such as the community manager and the IT manager.

Given the growing amount of digital surrogates, digital curators appreciate Creative Commons licenses because they greatly simplify the management of rights of use[11].

Digital curators are aware that cultural content aggregators have some usability issues (such as the way images are categorized and the limited search options). Thus, they share the collection through multiple channels in order to reach out to as many users as possible and collaborate to divulgative events organised by nonprofit organisations. For in-

Tab. 3.1
Needs and requirements of cultural collections.

	Need	Requirement
related to information need	Obtaining evidence of his work (mapping the status and use of the online collection)	To display status and spread of the collection
	Identifying any improper uses of the contents within the cultural content aggregator	To show the use of the collection
	Understanding which are the channels of diffusion of the collections and their characteristics	To show the channels of spread of the collection (within the cultural content aggregator)
	Quickly monitoring the collection as a whole and in detail	To easily switch from the overview to a detailed view of the collection
related to functional need	Loading content quickly	To allow batch upload
	Easily monitoring the collection as a whole and in detail	To easily switch from the overview of the collection to the details of a single item
related to non-functional need	Inserting content with annotations and exceptions	To allow the use of comments and annotations

	Need	Requirement
related to information need	-	-
related to functional need	Allowing the community to validate content Organizing contents in a flexible way Quickly accessing the content Allowing the community to enhance the classification	To edit content (for registered users) To organize content in a flexible way To navigate content in an easy way To edit the information system (for registered users)
related to non-functional need	-	-

Tab. 3.2
Needs and requirements of nonprofit organisations.

stance, ETH-Library not only released its digitised collection through Wikimedia Commons, but also through two digital archives: the Search Portal[12] and E-Pics Image Archive[13].
Digital curators use some tools to monitor the spread of the open collections. For instance, in Switzerland they mostly use Cassandra[14] and BaGLAMa 2[15].

Nonprofit organisations
Since they provide the means to access multiple digitized collections, their role is crucial. In fact, they define the collections to be included within the application (in most cases), the classification system and the modalities to interact with collections. Nonprofit organisations obtain funding from public or private institutions through the development of projects related to education and dissemination activities. They also receive donations from the cultural content aggregator end-users and, more generally, from those who believe and want to support their vision.
The choices of nonprofit organisations strongly impact on the actions of the other stakeholders. For this reason, usually every choice is shared and discussed with the others. Usually nonprofit organisations organize meetings with GLAM, communities, end-users and any other domain experts to discuss solutions to improve the cultural content aggregator in terms of both content and functionalities[16]. They also organize workshops with the involvement of the community to encourage the improvement and creation of content.The nonprofit organisations consists of a staff and an administration board that set and manage the cultural content aggregator strategy. The figures closely related to the digitised collections are the community manager, responsible for relations with cultural institutions, collections manager, content coordinator and web designer.

	Need	Requirement
related to information need	-	-
related to functional need	Easily reorganising content according to the dissemination activities Easily add and edit content	To reorganise digital surrogates To provide tools for creating and editing content
related to non-functional need	Creating sets of surrogates	To create sets of surrogates

Tab. 3.3
Needs and requirements of communities.

Communities

Communities consists of enthusiasts, nerds, teachers and other culture lovers who voluntarily contribute with the cultural content aggregator to improve or create content based on cultural collections. Usually volunteers, together with nonprofits, contribute to review the content provided by cultural institutions to make them more accessible for the general public. Within a community, volunteers choose the actions to perform in a shared and participatory way. Rewarding volunteers for their work the allows the increase of the spirit of collaboration and the generation of new content[17].

Within a cultural content aggregator, the community's activities consist in the creation, improvement and dissemination of content and can take place in multiple ways, both individually and collectively. Among the activities carried out individually there is the production of articles based on digitized collections. Collective activities include workshops, hackathons and talks and are mainly aimed at increasing public awareness of the collections. Writing challenges also appear to be good solutions for triggering volunteer participation and stimulating the writing of new Wikipedia articles or the improvement of very short poor-quality articles (Menichinelli et al., 2020).

Writers

Writers are people, like journalists, bloggers and editors of thematic websites, who publish articles on topics related to cultural collections. They use digital surrogates s graphical support to their both online and printed publications. The audience of content aggregators are experts of a specific domain, and in most cases, they have a clear idea of the item to be searched. However, in some cases, they ask for the support of communication agencies to search for images.

Writers generally prefer to get what they want quickly, they are unwilling to explore collections without real need.

	Need	Requirement
related to information need	Obtaining rich and structured information	To display multiple structured metadata
	Reading clear conditions of use of the surrogate	To clarify the conditions of use of the surrogate (including different licenses for different countries)
	Obtaining a clear idea of how the content are organised	To provide surrogates organised in a simple (or case easy to understand) structure
related to functional need	Obtaining a restricted and specific set of images of interest	To provide options to narrow the research or to filter results
	Comparing multiple items	To provide the possibility to compare two or more items
	Find items similar to the one selected	To show similar items based on metadata and image recognition (AI)
	Quickly accessing to good quality content	To provide download and sharing options
related to non-functional need	Analysing relationships among surrogates	To visualise connections among surrogates

Tab. 3.4
Needs and requirements of writers.

Communication agencies
Whitin communication agencies the people in close contact with digital collections are photographers and graphic designers. They usually work for writers and publishing houses as intermediaries in the search for images to be published. The final goal of photographers and graphic designers is to find images provided with relevant information, in high quality and with few usage restrictions. Because of time constrains the needs of photographers and graphic designers are closely linked to the functionality of the tools for accessing and using the collections. The work of communication agencies contributes to give visibility to cultural institutions and their cultural objects. Within the open cultural content ecosystem there is no real relationship between communication agencies and communities. However, communication agencies can benefit from the work they do.

3.3 Analysis of cultural content aggregators interface

In recent years, cultural collections have received more online visits then onsite visits (Borowiecki and Navarrete,

	Need	Requirement
related to information need	-	-
related to functional need	Obtaining a wide range of similar surrogates	To provide multiple similar surrogates (similar from the point of view of visual appearance and descriptive characteristics metadata)
	Creating customised categorisations of items	To provide a work area to organize items
	Minimizing the time for search	To provide tools that facilitate research operations
related to non-functional need	-	-

Tab. 3.5
Needs and requirements of communication agencies.

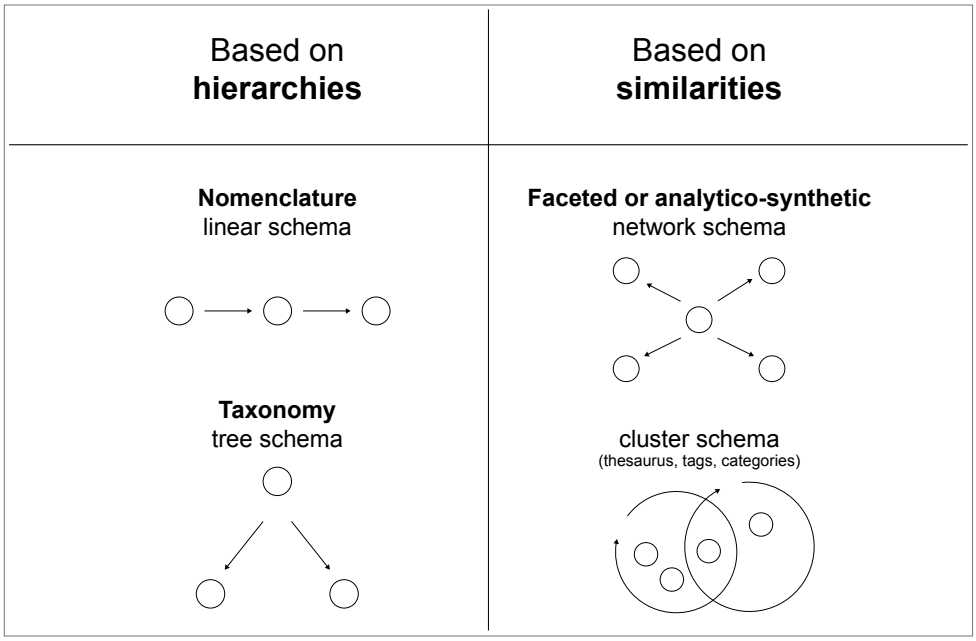
2016) and web browser represent the main channel of access, thus they are shaped according to the practices of interface design and information retrieval, as well as digitisation processes (Whitelaw, 2015). Accessing digitised collections via the Web is challenging both from the point of view of the content and of the user interface. The content should shift from the overall detailed description of the collection to item-by-item descriptions, while the user interface should foster the exploration of the collection as a whole and in detail. Many institutions have already developed the technology and the staff expertise necessary to cope with primary information challenges. Now, designers are involved in leveraging the vast array of digital surrogates in new, experimental and innovative ways (Deal, 2015).
The design of effective user interfaces requires the analysis of the motivations that push users to access and use digitised collections. Visits to museums and libraries have been associated with the information need and the fulfilling of recreational activities, such as curiosity and spending free time (Johnson and Thomas, 1998; Brida, 2015). The motivation for accessing online heritage fruition, instead, has been linked to remote access (Booth, 1998), but also to academic research, personal enjoyment, educational and commercial use and creative reuse (Borowiecki and Navarrete, 2016). Users contribute to online collections by enriching the cultural heritage material and filling resource gaps if the cultural institutions implement motivators such as money, interest in the topic, ease of entry and participation, altruism and meaningful contribution (Organisciak, 2010).
The choice of a specific classification system affects the design of the user interface, in particular, when the aggregator allows users to contribute to the classification. Europe-

an content aggregators mostly use a faceted classification system. The analysis of European cultural aggregators also showed that items metadata could be inconsistent and that "ungenerous" user interfaces make most of the items inaccessible.

3.3.1 Classifying the collection

The classification of digitised cultural objects consists of the addition of individual or multiple classes (or categories) to each item. The classification is essential for several reasons: first, it allows the retrieval of an item within a collection and facilitates the discovery of duplicates by aggregating items into specific classes. Moreover, it brings meaning and semantics. Classifying items is a time-consuming process but it also requires human, cognitive effort and computational load.Usability studies show that information seekers in extensive collections prefer meaningful groupings of related items, in order to quickly understand the relationships among them and to proceed with the exploration (Hearst, 2006). The choice of a classification system impacts the design of the user interface to access the items. It may also affect the design concerning the interactive tools needed to allow users to make and update the classification.

Fig. 3.4
Type of classification systems



Classification is not strictly related to a technological setup, but some databases are more suitable than others for collecting classified data, due to their models to organise data. For instance, records in a traditional database are organised in a treelike structure, while object-oriented databases store complex structures that are organised into hierarchical classes that may inherit classes (Manovich, 1999).

The classification should consider the uncertainty regarding cultural collection data, which is a relevant challenge in digitised cultural collections. Classification systems and user interfaces should make uncertainty visible by adopting design strategies able to deal with multiple interpretations of data (Windhager et al., 2019).

There are three main types of classification systems: nomenclature, taxonomy and faceted (or analytical-synthetic) (see fig. 3.4). Typologies describe, in abstract form, the relationships among classes. Each typology has his own classification schema describing, in visual form, the relationships among classes. There are four classification schemas: linear, tree, graph and cluster. The linear schema is based on ascending or descending order, the tree focuses on hierarchies, the graph on relationships, while the cluster is based on independent classes. European content aggregators mostly use facet classifications (see fig. 3.5).

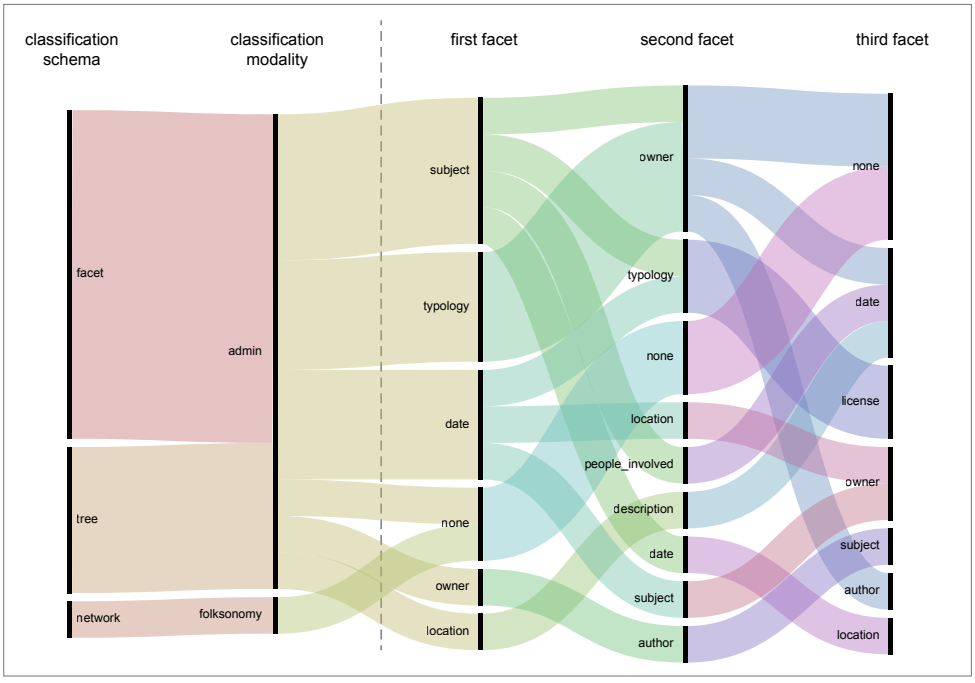


Fig. 3.5
Sankey diagram showing the access characteristics of a selection of 14 European cultural content aggregators.

Nomenclature (linear schema)
Nomenclature uses the linear schema. The classes are ordered according to a unique principle, such as the alphabetical order and the order of relevance. Within every class, items may be ordered according to the same principle or a different one. European cultural aggregators do not use nomenclatures. Some of them list items in alphabetical order. They list items in alphabetical or relevance order just when users do a search.

Taxonomy (tree schema)
Taxonomy uses the tree schema, where the classes follow a hierarchical structure and are nested within each other. Every class contains items that may follow a certain order or not. Items can belong to a single class in mono-hierarchies, or to multiple classes in poli-hierarchies (Rosati, 2006). In mono-hierarchies, digital surrogates are grouped in one, single collection, while in poli-hierarchies surrogates are grouped in multiple sets which are usually named collections, exhibitions or galleries.
An example of cultural aggregator based on taxonomy is the Swiss National Library[18] (see fig. 3.6), which gathers the bibliographic content of Switzerland. Its classification schema is a tree showing, at the first level, the partner cultural institutions and then, going towards deeper branches, more specific classes of digital surrogates. Some branches also consist of nomenclature of classes, ordered by alphabetical order of author but, from the analysis of the classification system, it emerges that this is not consistent. Within the classification system, every single cultural institution creates hierarchies according to different metadata, such as the medium, the collection and the city. Thus, the tree schema follows different principles of classification since the second level.

Facet (graph and cluster schema)
Faceted classification systems may use the graph or cluster schema. In the graph schema, classes are interconnected with each other in a direct, indirect or bidirectional way. In the cluster schema, every class is independent. The faceted classification has been introduced with the advent of the Web and the use of semi-independent or independent classes of items makes this system highly flexible and easy to update.
Faceted classification is the most widely used typology of classification system among European cultural aggregators.

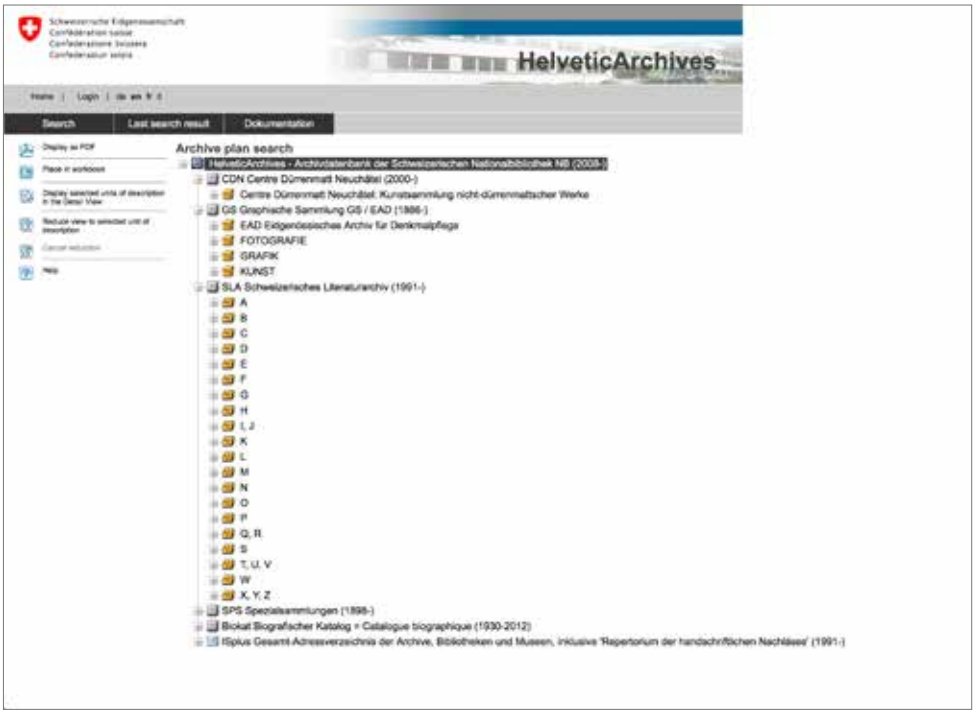


Fig. 3.6
Swiss National Library, treemap.

EUscreen[19], an aggregator of audio-visual heritage, uses a cluster schema (see fig. 3.7). The facet formula – the sequence of facets to produce convenient results while retrieving items – consists of two main steps. First, there is the medium facet (or class), containing the following sub-classes: video, image, audio, document, series/collection. Secondly, there are the following secondary classes: language, decade, topic, provider, genre, country of production and publisher.

Wikimedia Commons – the multimedia library of Wikipedia (see fig. 3.8) – uses a graph schema. Within the content aggregator funded by Wikimedia Foundation, categories (or classes) can be nested inside each other, allowing the creation of direct, indirect and bidirectional relationships among categories. The classification system is not coherent: it consists of categories characterised by very different level of specificity. For instance, the category "Radio" points to the categories "History of radio" and "Asociación Latinoamericana de Educación Radiofónica". Furthermore, the user interface displays it as a taxonomy. These issues make the system difficult to comprehend and to navigate.

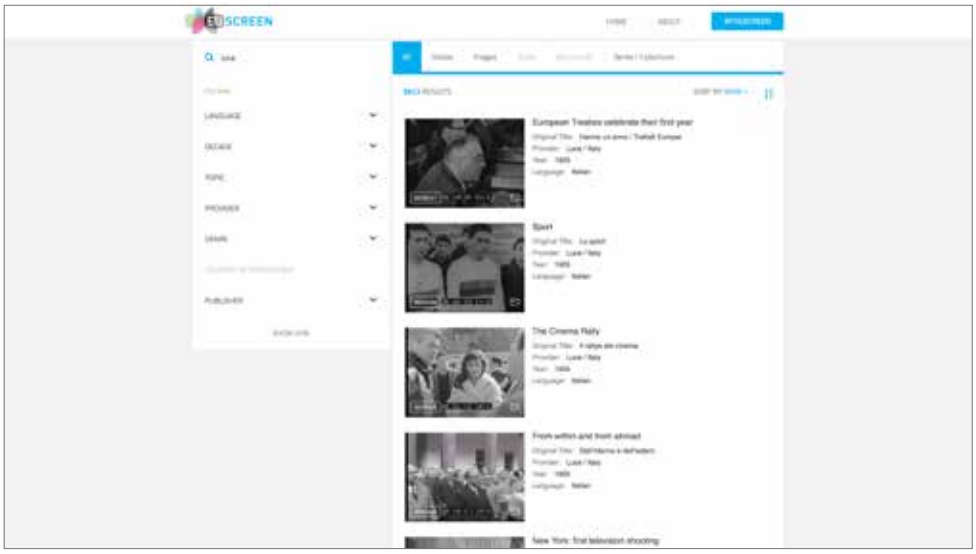


Fig. 3.7
Euscreen, facet.

Classification systems are based on the characteristics of the items constituting the collection and he classification grounds on the analysis of the metadata of every item (Zeng and Jian, 2008). Classifying items is a process that can be performed both manually by humans and in an automatic way by computers.

Human classification

Classification performed by humans is a time-consuming process. Introduced in ancient times to classify and make bibliographic resources easier to access, the outcomes of human classification are nomenclatures, taxonomies and facets. They are developed according to descriptive, structural and administrative metadata. Nomenclatures mostly use descriptive metadata (such as author name or date), while taxonomies mostly use descriptive and administrative metadata, such as the institution owner and the city of provenance. Facets, finally, generally use the three types of metadata, including the medium (such as images and videos).

In European cultural aggregators, administrators chose the classification system and its classes, while cultural institutions provide the digital surrogates and related metadata. Administrators usually chose the classes according to the set of mediums.

Almost every European aggregator use a human classification made by administrators. Wikimedia Commons, instead, uses a folksonomy[20] - a collaborative classification process

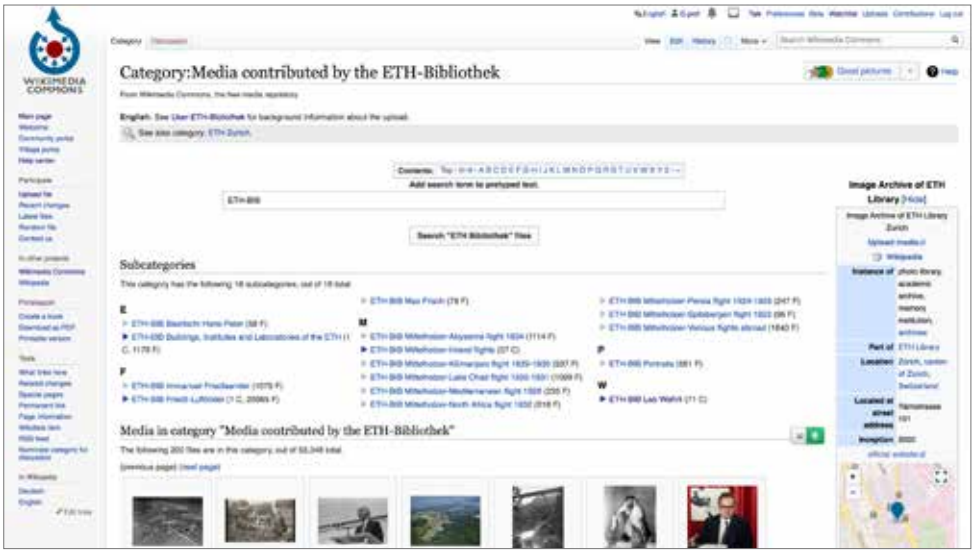


Fig. 3.8
Wikimedia Commons, category page.

made by end-users. Folksonomy offers multiple advantages to users: the possibility to organise personal content, enable pivot browsing through a re-grouping of the content and to enable the user to discover similar contents and users based on shared tags (Stiller, 2014).
Folksonomy only works with facets in the form of tags or categories. The use of these independent or semi-independent classes does not require a global knowledge of the classification system. On the Internet, several applications, such as Pinterest and Good Reads, use folksonomies.
Admin classification and folksonomy have both pros and cons. On one hand, admin classification may guarantee the consistency of the system, but it requires plenty of time and effort by a small group of people. On the other hand, folksonomy is easy to perform by several people but it may generate inconsistency in terms of classes specificity.

Computer classification

The classification performed by computers has been introduced in recent years, following the technological advancement in the field of computer science. It is possible thanks to the use of several technologies, such as neural networks, statistical algorithms, machine learning and artificial intelligence. The computer automatically classifies items according to characteristics that humans may find difficult to evaluate (such as the visual similarity among items) or boring and repetitive (such as the file format, and the ori-

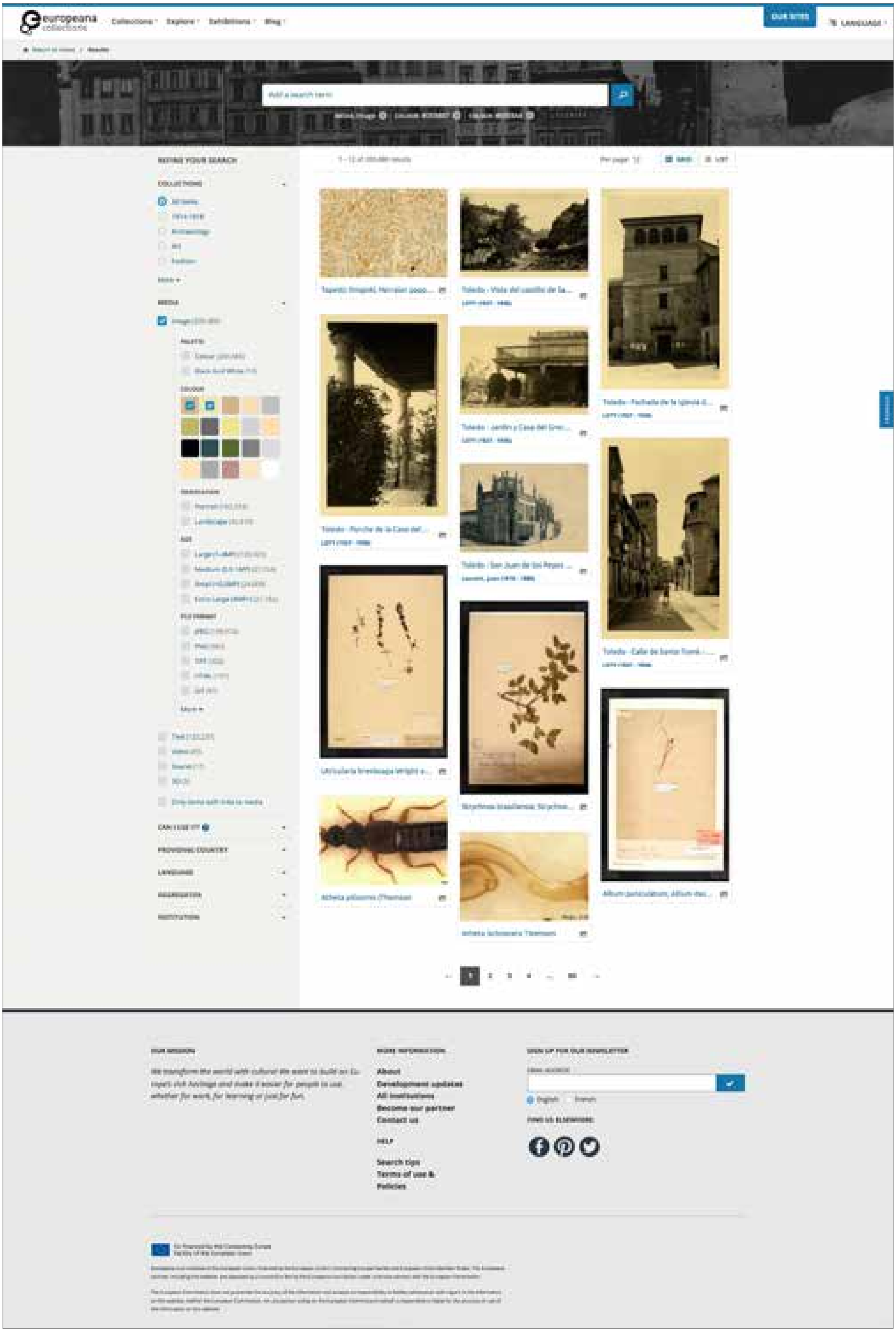
entation). The outcomes of the computer classification are facets, which mainly use structural metadata, such as the colour palette and the visual structure. In contrast with the univocal results given by the human classification, the computer classification returns results with a certain percentage of confidence.
Among European content aggregators, Europeana and Wikimedia Commons apply advanced computer classification. Europeana (see fig. 3.9) classifies items by colour palette and this classification process generates a six-colour palette (consisting of both the name and the hexadecimal colour code) for every item. Wikimedia Commons automatically applies classes to an item under the form of hidden categories, which mostly refer to administrative metadata, for instance, "CC-BY-SA-4.0" and "Self-published work".

3.3.2 Accessing the collection

The growing volume of contents combined with the pressures of time and money makes the need to improve items findability[21] a critical issue (Morville, 2005). In order to reach this goal, user interfaces of every European cultural content aggregators use at least two tools to allow end-users to access digital surrogates. The homepage generally provides all the tools to explore the collection, while the internal pages provide both tools to narrow and expand the search (see fig. 3.10).
From the literature review and the analysis of cultural content aggregators, it emerges that there are three access modes: search, browse and explore. Every access mode has his own tools - interactive interface elements that allow the user to access the collection - and can be applied in any classification system. However, access modes suit best for specific classification systems. In history, information systems have introduced the search, browse and explore access modes sequentially and tools belonging to the three different modes can coexist.

Search
The search access mode was introduced in the 1980s. It refers to a search engine which allows users to ask for information through the submission of a query. From the user interface point of view, search mode may consist of both simple and advanced search boxes. A simple search box allows users to search for information among all the availa-

Fig. 3.9
Europeana, colour palette facet.


europeana collections
OUR SITES
LANGUAGE
REFINE YOUR SEARCH
We transform the world with culture! We want to build on Europe's rich heritage and make it easier for people to use, whether for work, for learning or just for fun.
About
Development updates
All institutions
Become our partner
Contact us
Search tips
Terms of use & Policies

ble content, while an advanced search box allows users to search for a piece of information among specific contents or areas of the information system. It also may allow the use of logic operators[22].

Search crosses the whole set of surrogates' metadata. Thus, it can be applied to every classification system. However, the fact that the search process generates several results requiring a further filtering operation makes the search mode more suitable for faceted classification systems.

Search is an access mode which requires a basic knowledge of the information system or the content domain to work properly. It suits best users who already know what to look for.

Every European cultural aggregator allows users to access the collection via search mode. Culture Grid[23] - an English aggregator funded by Collections Trust - provides users with one of the most advanced search boxes (see fig. 3.11). The advanced search takes place through two or more of the following steps: the submission of a query, the selection of metadata where to focus the research (such as in the title and the place), the choice of the medium and the selection of a specific collection. Culture Grid also offers a search box

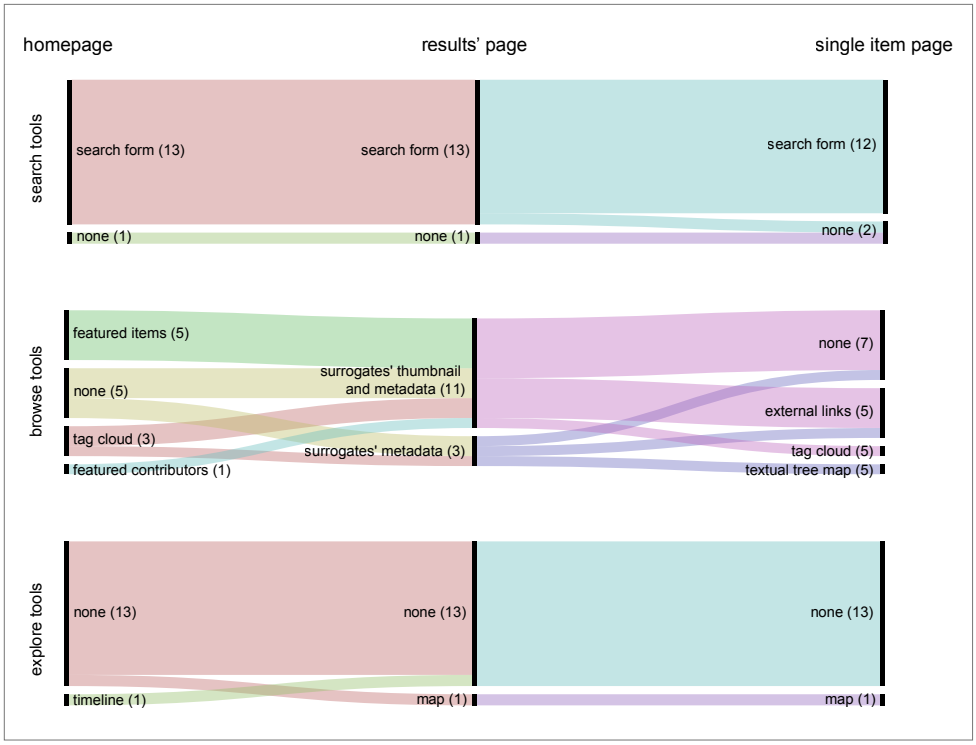


Fig. 3.10
Sankey diagrams showing the access tools of a selection of 14 European cultural content aggregators.

builder - a user interface to build a custom search box that can be added to external websites.

Browse

The browse access mode was introduced in the 1980s, as well as the search mode. Browse consists of several tools, such as tag clouds, tree chart and list of categories that allow users to navigate among all the available content. There are two main typologies of tools for browsing the collection: the tools for accessing the collection and the tools for continuing the navigation.

The first ones, such as the list of categories (usually presented through image thumbnails), are available in a homepage and feature items selected by aggregator administrators, allowing the user to get an overall idea of the collection content. The tools for continuing the navigation, instead, allow the user to keep browsing items starting from the content of his interest. Among these tools, there are lists of related items (mostly based on the same author and category) and a portion of the collection tree.

The browse mode allows accessing classification schemas with multiple levels of nesting quickly. Due to this, browsing tools are very useful in taxonomies having both mono-hierarchies and poli-hierarchies.

Browse is an access mode that requires time to reach a potentially interesting content and it is more suitable for users who do not have a specific goal to accomplish.

Interface elements for browsing collections are widely used on cultural content aggregators. An example of European content aggregator using browsing tools is the European Film Gateway (EFG)[24] (see fig. 3.12), which displays historical documents coming from European film archives and film libraries, such as photos, posters, programmes and censorship documents. The aggregator features in the homepage the recent uploads, the popular items (divided according to medium and period of popularity) and the video of the day. Within a page in the primary level of the information architecture, EFG features a list of all the collections.

Explore

Cultural institutions have applied data visualisations to digital collections since the 1990s. At that time, researchers started to theorise the concept of "Visual Information Seeking" (VIS) as the use of visual displays for designing systems to facilitate the visual method of search. The VIS

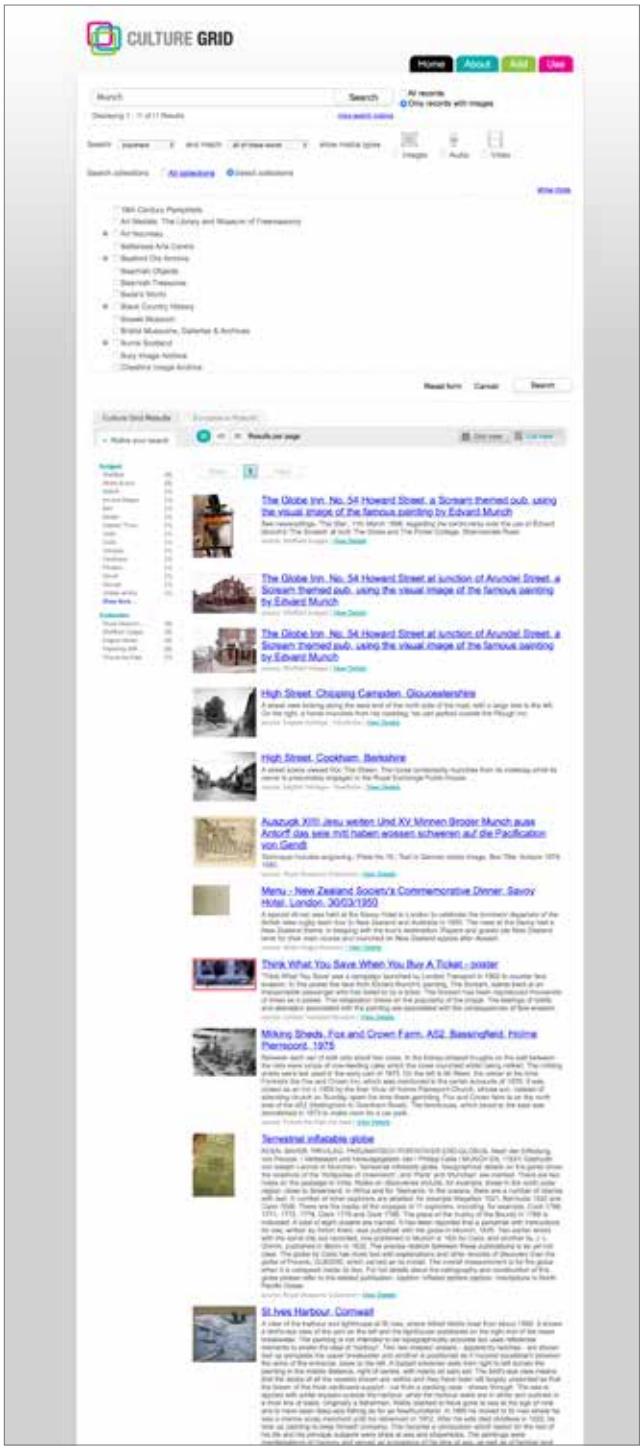


Fig. 3.11
Culture Grid, advanced search box.

mantra says: "Overview first, zoom and filter, then details on demand" (Shneiderman, 1996).
Despite these studies, visualisation has not become a common feature of most digital archives (Deal, 2015).
At the core of the explore mode, there is the idea to provide an overview of the collection first, and then present the items detail: this makes the explore mode a good solution to analyse the collection as a whole. By elaborating every metadata independently from the rest of the object metadata, the explore mode visually shows collection features such as all authors, places and dates.
This idea behind the explore mode is in contrast with the one applied by search and browse modes, which mainly focus on the single surrogate. In fact, by elaborating objects metadata as a whole, search and browse modes quickly return lists of items. However, they may need further search

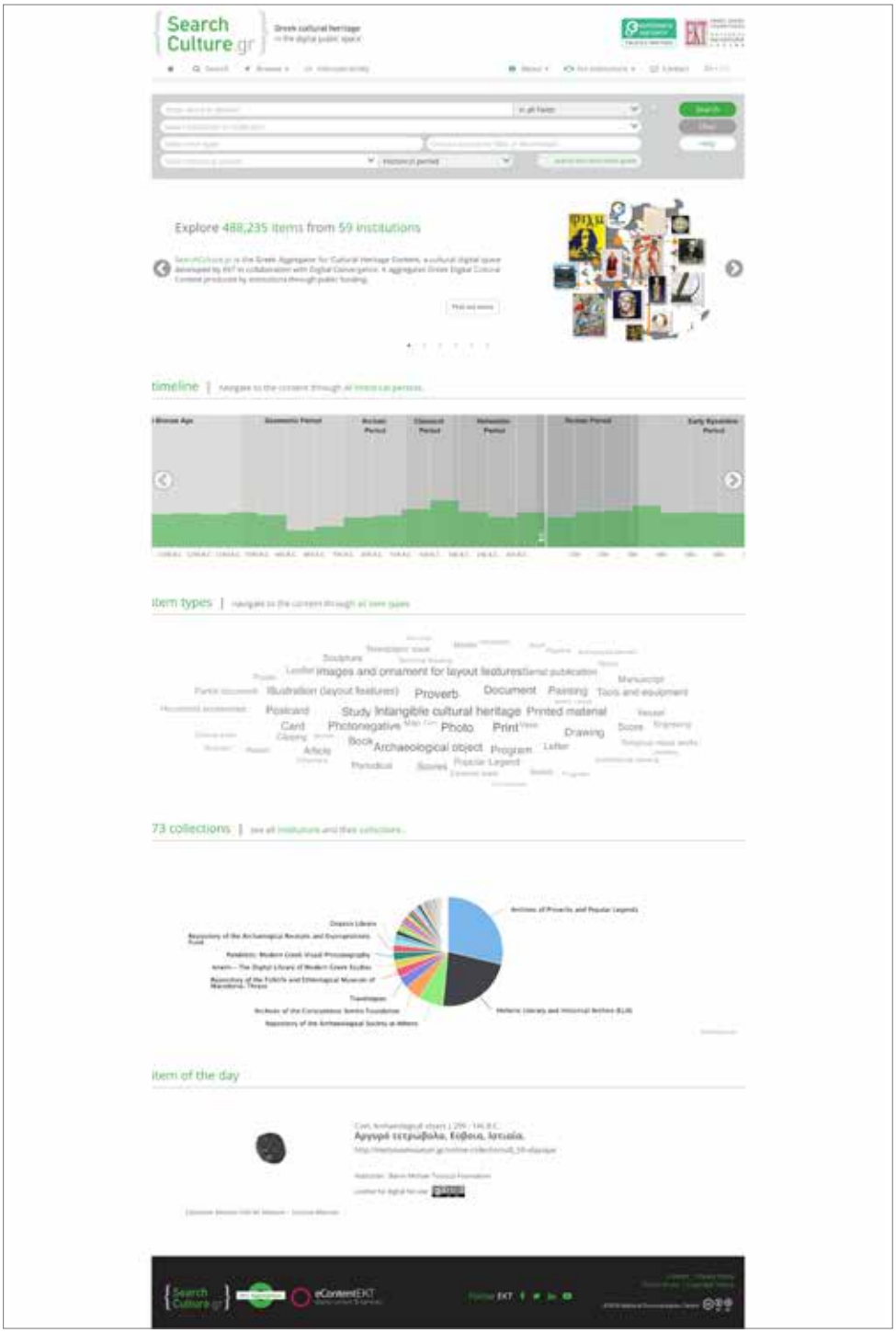
Fig. 3.12
The European Film Gateway, popular items.

and analysis by the users.
The explore mode mainly uses 2D and 3D visualisations, as well as visual filters as visual tools to analyse and access the collection. The visualisations display objects metadata. Since one single visualisation of the collection is not enough to explore every collection dimension (Dörk et al., 2017), the "coordinated multiple views" method has been established: it is also called "the parallax effect" (Drucker, 2013) and consists of multiple data visualisations linked one another through exchanging references to selected items (Andrienko et al., 2007).
From the literature, it emerges that digital archives use temporal and non-temporal visual models. Temporal visual models essentially include the timeline (in one, two or three dimensions), while non-temporal visual models include maps, networks and plots (Windhager et al., 2019). While, in digital archives, several visual models are widely used, European cultural content aggregators adopt only two visual models: timelines and diagrams. A timeline visually represents objects metadata organised by a model of time. It can allow users both to extract new knowledge from existing digital collections and present new findings (Kräutli, 2016). The temporal arrangement of collections helps users to identify meaningful connections between items (Ruecker et al., 2011).
The most used visual models are timeline, network and bar charts, which summarise objects metadata in order to provide an overview of the content. The timeline is among the earliest examples of visual renderings of data and enables the study of temporal relationships within a cultural collection (Kräutly, 2016). The network diagrams depict various relationships among cultural objects. They tend to be analytical, but they can also be used for general exploration (Dörk et al., 2017; Herseni et al., 2018). The bar chart, instead, shows distribution patterns at a glance (Whitelaw, 2015).
Even though collection visualisations keep visual displays separated by digital surrogates, a study has shown that a promising direction of work is exploring new ways to combine collection metadata with the material texture of collection items (Glinka et al., 2017).
The explore mode is primarily based on aesthetic qualities of collection and interaction design. Users can explore the collection through direct and procedural manipulations, which is the use of interactive elements that immediately

and continuously display results according to user demand (Ahlberg and Shneiderman, 1994). Procedural manipulation refers to the adoption of a linear process of getting results, where every step supports specific operations (Botta et al., 2006).
The use of interactive visualisations allows users to gain an overall idea of the collection, but it may need to be integrated with search and browse tools.
The explore mode can be applied to every classification system. However, its ability to display multiple collection dimensions visually, even within limited space, make it more suitable for faceted information systems with several facets.
Among European content aggregators, the one who widely uses the explore mode is Search Culture[25] (see fig. 3.13): it displays a timeline, a tag cloud and a pie chart showing the number of items for a cultural institution in the homepage. The timeline is the most relevant visualisation because it groups items by century and makes the various historical periods, from the early Bronze Age to modern times, visible. By clicking on every interactive element of visualisations, the user can reach a page with the results of his selection.

3.3.3 Navigating the collection

The goal of a cultural content is not just to provide the end-user with an access to a specific digital surrogate of interest, but also allow him to navigate the rest of the collections easily. European cultural content aggregators often lack of easy-to-use tools for navigating the collection. However, all of them adopt a very simple information architecture[26], which consists of three main pages: the homepage, featuring some digital surrogates, the list of results requested by the user and the detail page of the selected surrogate. The three pages are interconnected and the direct link between the detail page and results' page is missing only in some aggregators.
The homepage generally provides all the tools to explore the collection, the results' page allows to narrow the exploration and the detail page gives the possibility to continue exploration.
From the benchmark of the European cultural content aggregators, we can identify four typologies of navigation tools: for navigating among pages, for navigating within a page, the classification system and the related content (see fig. 3.14).


Search Culture.gr
Explore 488,235 items from 59 institutions
timeline
item types
Sculpture
images and ornament for layout features
Illustration (layout features)
Proverb
Document
Painting
Tools and equipment
Postcard
Study
Intangible cultural heritage
Printed material
Photonegative
Photo
Print
Drawing
Score
Card
Book
Archaeological object
Program
Letter
Article
Periodical
Scores
Popular Legend
73 collections
item of the day
Αργυρά τετράβολα, Εύβοια, Ιστιαία.
eContentEKT

Tools for navigating among pages

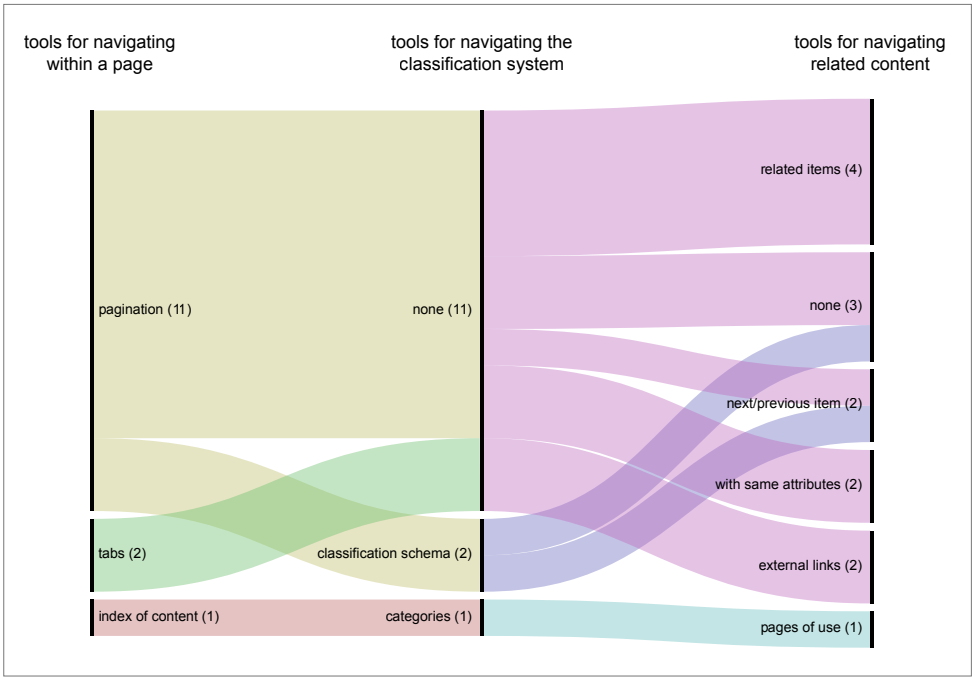
Tools for navigating among pages are the conventional web navigation tools. They allow end-users to understand where they are within the information architecture, where they can go and where they come from. Among these navigation tools there are the menu, generally positioned at the top of every page, the arrows to go back to the previous page and the breadcrumbs, a secondary navigation tool showing the navigation path from the homepage to a specific internal page.

Fig. 3.13
SearchCulture, catalogue visualisation.

Tools for navigating within a page

Tools for navigating within a page are mainly used within the results page: they allow end-users to rearrange digital surrogates and to obtain different overviews. Among the internal navigation tools there are the pagination - an interface element that splits the retrieved digital surrogates into discrete pages, widely used in websites - and the list/grid dropdown menu - an interface element that allows end-users to choose whether to see results as a list or as a grid.
Due to the big amount of results ranked in the results page, the sorting options are an important tool to rearrange the content easily. The most used sorting options are in alpha-

Fig. 3.14
Sankey diagram showing the navigation tools of a selection of 14 European cultural content aggregators. In the cases in which cultural content aggregators feature several tools belonging to the same typology, only the most prominent is relieved.



betical order of title, author or ordered according to the creation date.
Sorting and pagination tools are important because they prevent end-users from being overwhelmed by a mass of contents in one page. Thus, the content aggregator audience can focus on small sets of surrogates and metadata.

The National Library of Finland[27] features interesting navigation tools within the results page (see fig. 3.15). It provides end-users with the opportunity to sort results by relevance, date, typology, author, title and last modified or added. It also allows users to select how many results they want to display per page (20, 50 or 100) and in which form (grid, expanded list and condensed list). The user interface also featuresa pagination at the top of the page. Users interested in viewing a larger preview of the digital surrogates can click on the thumbnail to open a full screen media viewer.

Tools for navigating the classification system
Tools for navigating the classification system refer to interactive elements that allow the end-user to navigate the classification schema of the cultural content aggregator visually. These tools essentially consist of the classification tree

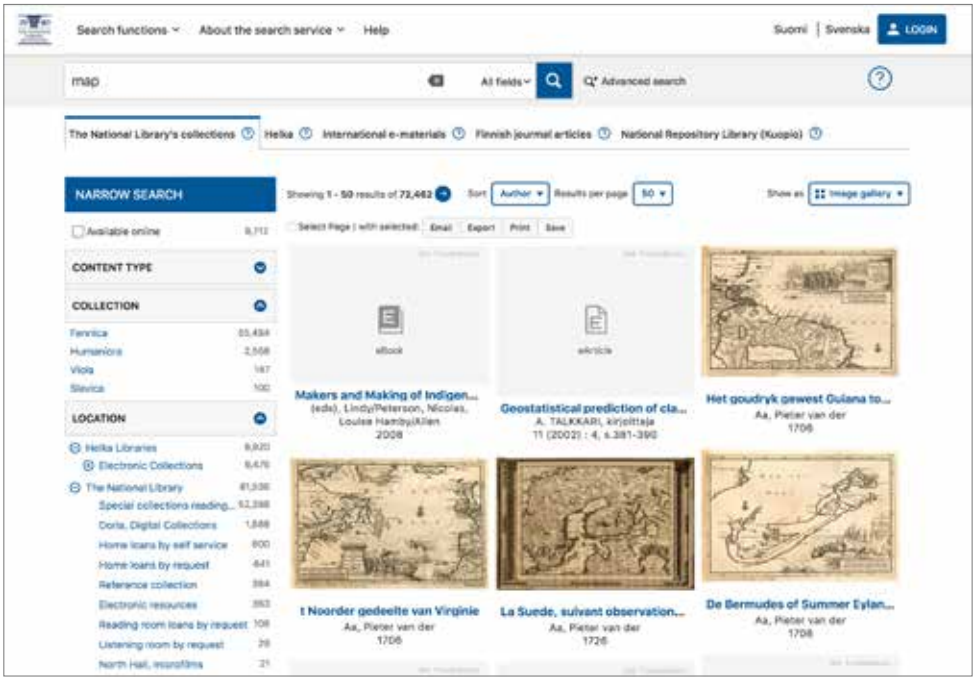


Fig. 3.15
National Library of Finland, tools for navigating the results' page.

and the categories (as a list or as a tag cloud). Some cultural content aggregators also feature the map of the GLAMs contributing to the web platform as a way to navigate the available collections geographically.
An example of cultural content aggregator using tools for navigating the classification system is the Swiss National Library (see fig. 3.16), which provides users with a tree schema as a visualisation of the classification system. The representation is placed in every page containing the single item. It shows the hierarchy related to the item and all the other elements belonging to the same branch. The tree schema uses different icons to distinguish between GLAms, collections and items. This navigation tool is useful for

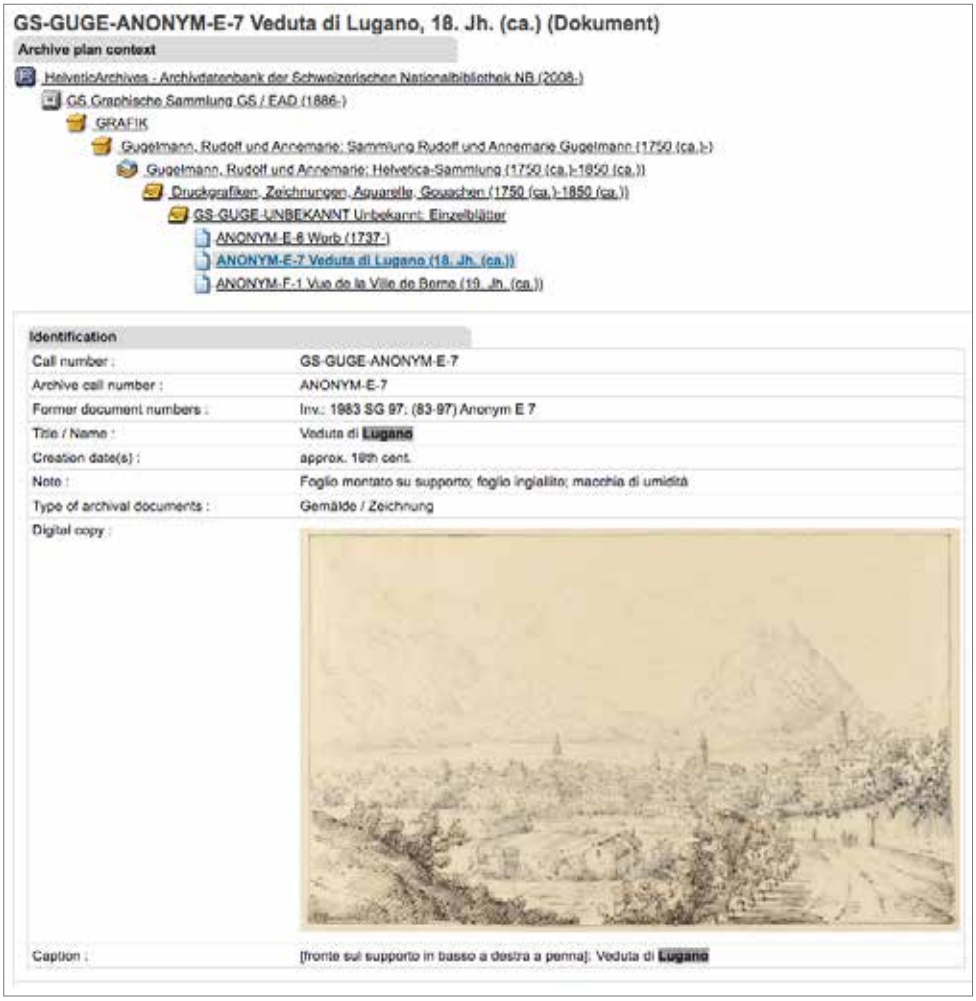


Fig. 3.16
Swiss National Library, classification schema within an item page.

end-users who wish to navigate the digitised collections without a specific search scope.

Tools for navigating related content
Tools for navigating related content allow end-users to continue the navigation when they reach a specific item. Among these tools there are the panel with related surrogates, the panel for the visualisation of surrogates with an attribute in common with the selected item and the panel with a list of external links.
An example of cultural content aggregator using tools for navigating related content is Europeana (see fig. 3.17) which features, in the item page, some panels showing the surrogate metadata. These metadata link to pages containing a list of surrogates with the same metadata.

3.3.4 Using the collection

The aim of a cultural content aggregators is to provide users with meaningful, high-quality content and digital surrogates to be used in their studies and publications. While some European cultural content aggregators only provide access to collections, some other adopt tools to facilitate the use of digital surrogates.
User can access these tools in the page containing the single

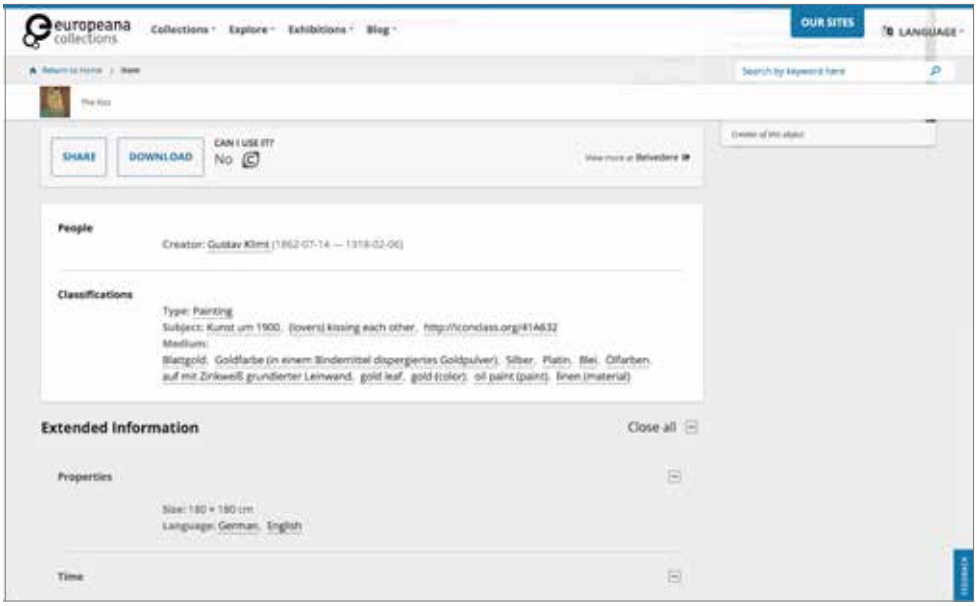


Fig. 3.17
Europeana, links to surrogates with similar properties.

image. The most basic ones are available for everyone, while the advanced ones are accessible only after the creation of a personal account within the content aggregator. From the benchmark of the European cultural content aggregators, we can identify four typologies of tools that allow the usage and the spread of digital surrogates: editing, organising, generating and sharing tools (see fig. 3.18). While the tools for navigating the collections are mainly conceived for browsing and exploratory purposes, the tools for using the collections are mainly thought for elaborating content.

Tools for editing content

Tools for editing content enable users to edit items' metadata or propose an edit directly. They consist of online forms provided with one or multiple fields that users can fill in with more accurate or missing information. The use of these tools allows a constant collective update and improvement of the content. However, the possibility to edit the content directly may generate

Wikimedia Commons (see fig. 3.19) is the most known content aggregator that adopts editing tools is. Both registered and unregistered users can edit images metadata from the edit panel. The editor displays content both in visual form

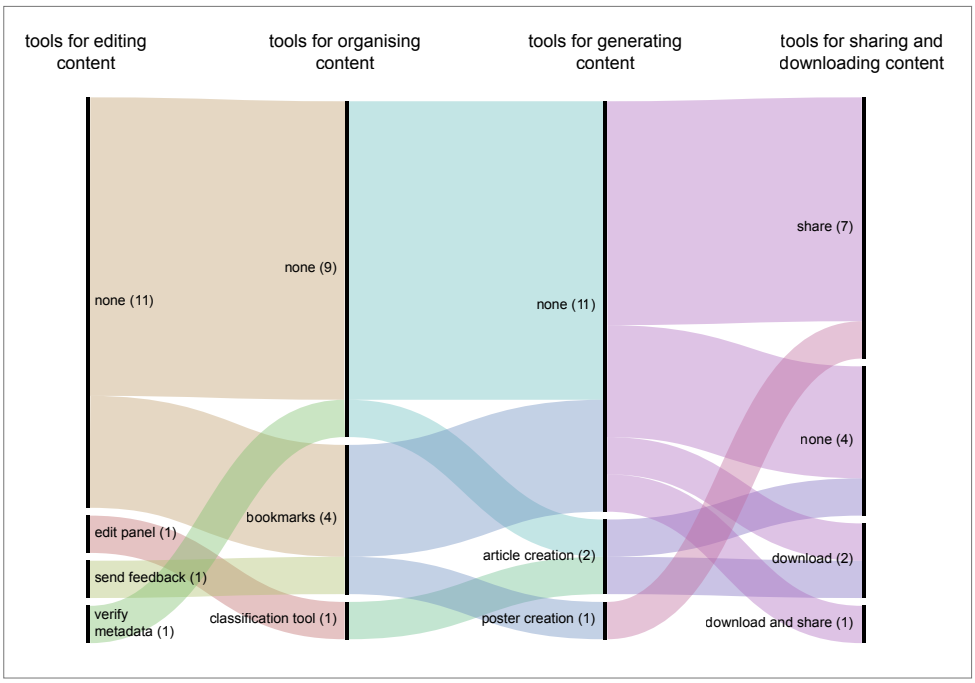


Fig. 3.18
Sankey diagram showing the tools for using digital surrogates of a selection of 14 European cultural content aggregators.

Fig. 3.19
Wikimedia Commons edit panel.

and as source code (in Wiki format) and allows users to edit all image metadata except for some hidden categories.

Tools for organising content

Tools for organising content allow users to group or save items of interest. These tools can serve as classification or bookmark tools. Classification tools refer to interactive elements that allow the creation of a folksonomy. Typically, these tools consist of an interactive text area, below the digital surrogate, where user can add his tags or categories. The use of classification tools allows users to add further information and discover similar content. Bookmark tools refer to interactive elements that allow the creation of personal collections of items. Aggregators use these tools in the form of bookmarks, favourites, watch list and collections.

Classification tools have an impact on all the users accessing the aggregator, while bookmark tools usually generate a group of items which is only visible to individual users.

EUscreen adopts bookmark tools (see fig. 3.20). Within the item page, a button allows the user to add the item to a custom collection or the list of bookmarks quickly. The saved item is then displayed in the MyEUscreen page, the user profile, which give users the possibility to sort items by date or by name. Within MyEUscreen page, a user can also create

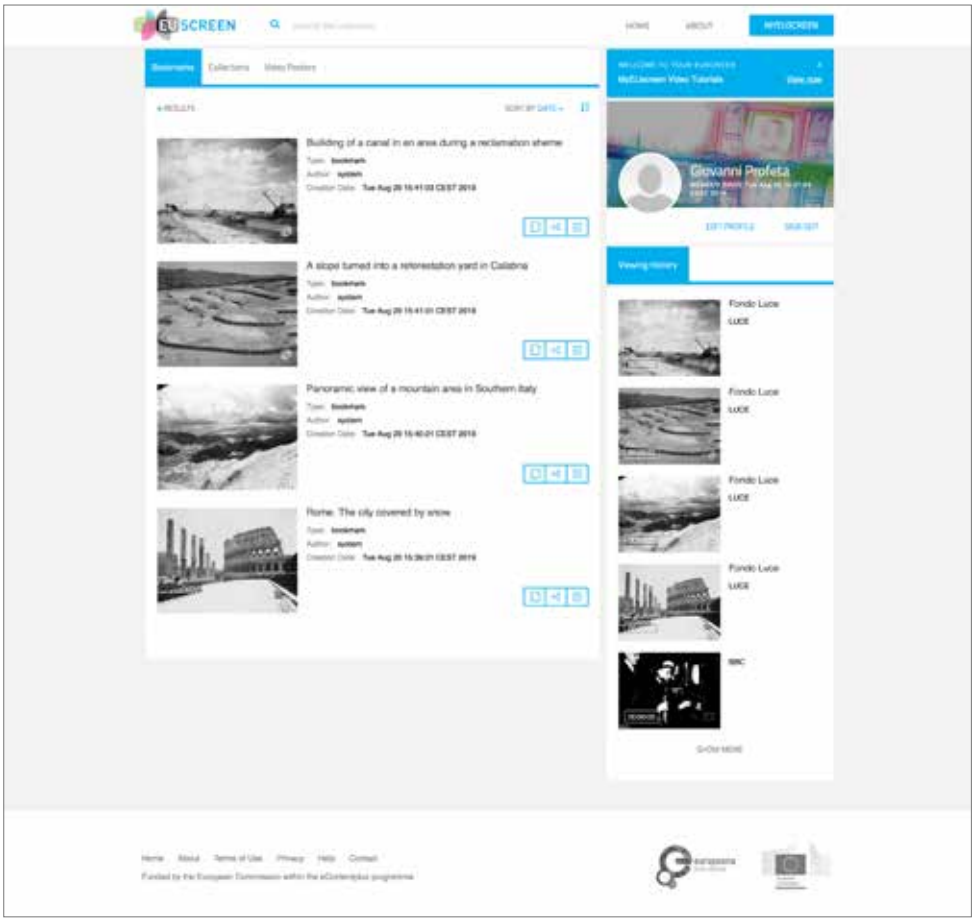


Fig. 3.20
MyEUscreen, bookmarks

video posters, an interactive panel with multiple video players that can be embedded in external websites.
Furthermore, MyEUscreen offers the Viewing History, that isthe list of the recent items viewed by the user.

Tools for generating content

Tools for generating content refers to the opportunity to generate, within the content aggregators, articles based on the digitised collections. User-generated content does not only enhance the access to cultural collections, but also foster the discovery and the investigation of new topics.
Among European content aggregators that allow users to generate content there is Kultur Pool, an Austrian portal to arts, culture and education initiative. It offers the users the opportunity to write and access "Smartworks", which

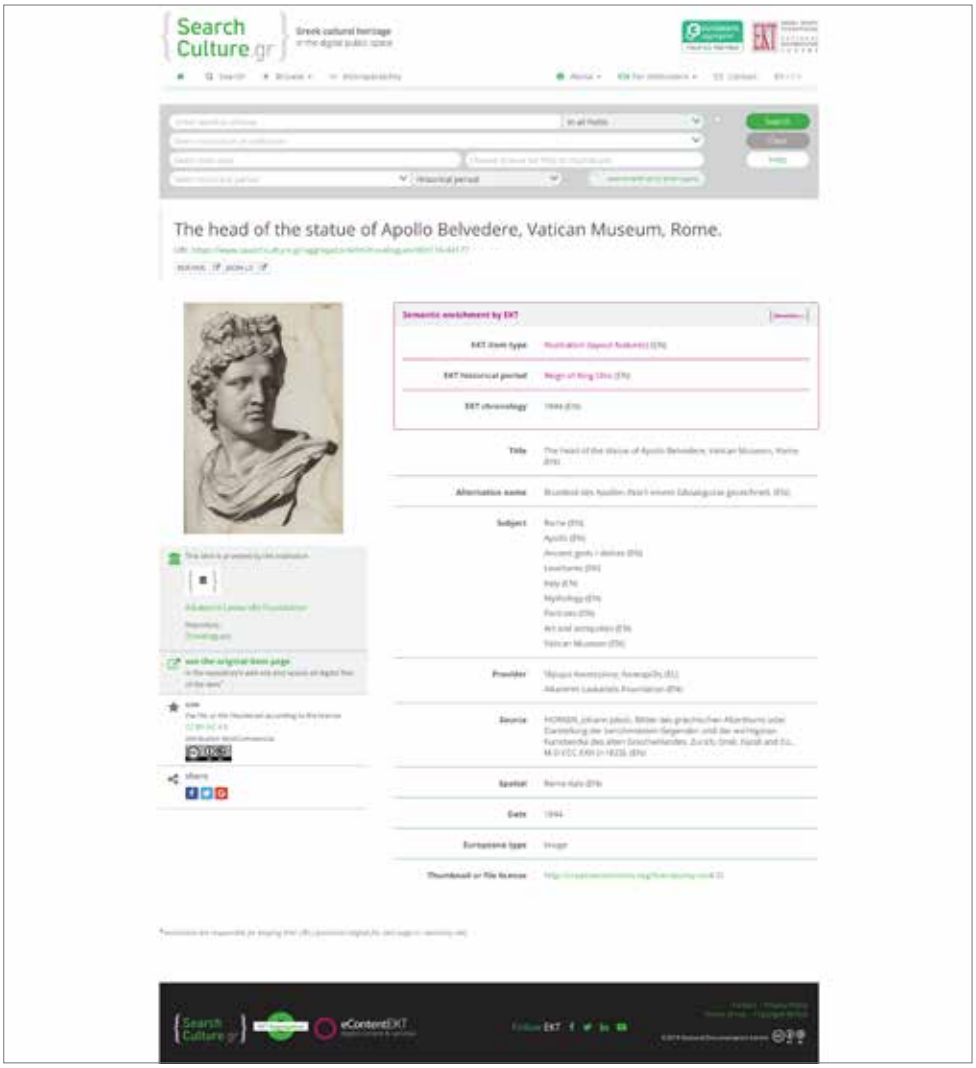
are essays mostly focusing on relevant artists who feature images coming from multiple cultural institutions (see fig. 3.21). Smartworks are featured in homepage and searchable through a search box and list.

Tools for sharing and downloading content

Tools for sharing and downloading content are the essential tools an aggregator can offer. They consist of panels containing links for sharing an item on social networks or other external websites and download it in multiple sizes. Aggregators which also contain bibliographic items may provide users with citation information and with the text-based file format.

Search Culture, for example, displays buttons to share content on Facebook and Twitter, as well as buttons to down-

Fig. 3.21
Kultur Pool, user-generated content.



load items metadata in XML and JSON format (see fig. 3.22).

Fig. 3.22
SearchCulture, download options.

Tools for monitoring content

Cultural content aggregators also offer visual tools that allow cultural institutions to monitor the access and the usage of the collections. These tools are based on usage data (or paradata), such as the number of visualisations, where the surrogate is used within the aggregator and each user.
The visualisation of usage data is very relevant for cultural institutions because it allows to understand the visibili-

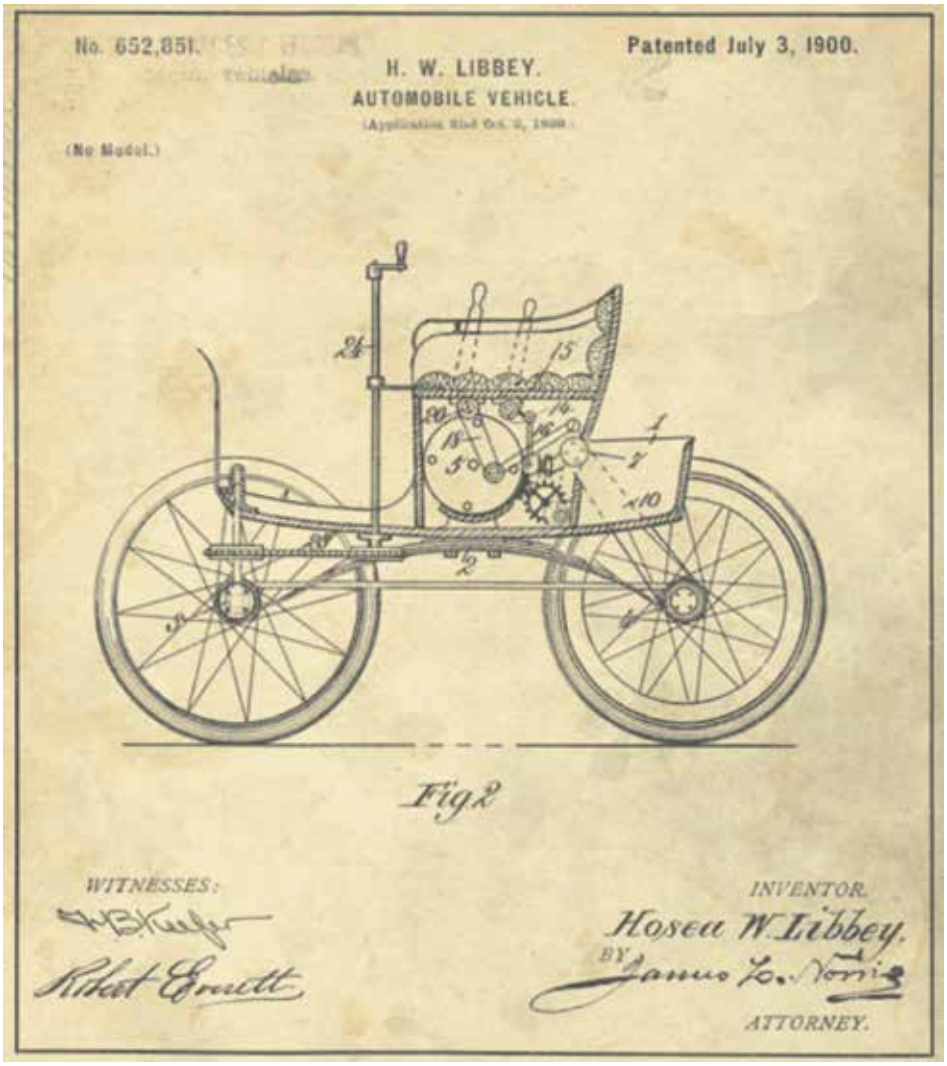
Fig. 3.23
BaGLAMa 2, Magnus Manske, 2014.

ty and the impact gained by the digitised collection. It also may provide insights concerning how to foster access to singles digital surrogate.

An example of tool for monitoring content is BaGLAMa 2[28] (see fig. 3.23), which shows a list of Wikimedia Common categories. Cultural institutions can see their categories aggregated by a certain period and by linguistic version. The details page may link to other Wikimedia tools. All the information about the linguistic versions and the overall page views are presented through tables, while he page views are displayed through a line chart.

Notes

1. LIFE Tags website: https://artsexperiments.withgoogle.com/lifetags/.
2. LIFE magazine was the most popular photojournalism magazine in the United States.
3. For "LIFE Tags" Google used its Image Content-based Annotation (ICA) algorithm to generate labels based on image pixels.
4. The "long tail" refers to the proportion of the products representing the best sellers and the rest of the products (also called "niche markets"). In traditional market the proportion is 20-80%, while on the Internet niches have a higher percentage of the market.
5. Wikimedia Commons guidelines: https://commons.wikimedia.org/wiki/Commons:Policies_and_guidelines.
6. Visual model refers to archetypes of charts, such as the bar chart and the box plot.
7. A feedforward is a form of hinting which provides users with clues about the effect of an action before it takes place.
8. Information gathered from Michael Gasser, head archives of ETH-Library.
9. A stakeholder requirement is a statement provided by a stakeholder on what he needs to solve a specific problem or respond to a specific need.
10. The paragraph presents information gathered from Michael Gasser, head archives of ETH-Library.
11. Cultural institutions usually release under open licenses only a small part of the digitised collection. For instance, ETH-Library has released about 5% of its collection.
12. ETH-Library Search Portal: https://www.library.ethz.ch/en/.
13. ETH-Library e-pics: https://www.e-pics.ethz.ch/en/home_en/.
14. Cassandra website: https://stats.wikimedia.swiss/
15. BaGLAMa 2 website: https://tools.wmflabs.org/glamtools/baglama2/index.html.
16. Information gathered from Florence Devouard, chair of the Wikimedia Foundation (2006-2008).
17. Information gathered from Dario Crespi, coordinator of Lombardia Region for Wikimedia Italy.
18. Swiss National Library website: https://www.helveticarchives.ch.
19. EUscreen website: http://euscreen.eu.
20. Folksonomy, also called social tagging, was coined by the information architect Vander Wal by binding together the two terms "folk" and "taxonomy". Folksonomy produces user-generated metadata.
21. Findability is the ease with which information or content contained on an information system can be found.
22. Logic operators are symbols used to connect two or more words in a search query. They allow users to make advanced queries, such as searching for a specific sentence and exclude a world from the search results.
23. Culture Grid website: http://www.culturegrid.org.uk/.
24. European Film Gateway website: https://www.europeanfilmgateway.eu/.
25. Search Culture website: https://www.searchculture.gr/.
26. Information architecture refers to the organisation of information in an information system so that it fosters a good user experience.
27. National Library of Finland website: https://kansalliskirjasto.finna.fi/.
28. BaGLAMa 2 website: https://tools.wmflabs.org/glamtools/baglama2/.



Ghosts Carriage,
by by Franck Grosbois, using this drawing from the National Archives and Records Administration, GIF IT UP contest, 2017. Website: https://gifitup.net/

"GIF IT UP" is an annual gif-making contest for the creative reuse of digitised cultural collections. The contest is run by Europeana in cooperation with Digital Public Library of America, Digital NZ and Trove. Every participant is asked to create a gif picking a copyright-free or an open licensed item from Europeana collections or other partner institutions. The prizes categories also include the first-time GIF-maker and the Industrial Heritage Category. All eligible entries are released under a Creative Commons license and showcased on the GIPHY channel dedicated to the contest.

4

Design projects

Cultural content aggregators are relevant because they give the opportunity the opportunity to find open digitised items coming from multiple digital archives in one single platform. However, the research has shown that the user interfaces of these web platforms have some usability issues, such as lacking and unclear content, misleading classification and limited access tools. Thus, most of the available heritage is invisible to end-users.
The design project generally aims to identify design solutions to solve the existing usability issues and to test some interface solutions that could facilitate the access and use of cultural collections. The results gained by the design project contribute to the definition of a set of design guidelines.
The design project consists of two different projects "Map the GLAM" and "GLAM Culture Hub". Map the GLAM, developed between February 2017 and August 2018, is a visualisation project exploring a cultural collection within Wikipedia. GLAM Culture Hub, instead, was developed between May-July 2019and it is an interactive prototype of a cultural content aggregator.

Map the GLAM consists of a set of data visualisation which explores the status and the diffusion of the ETH-Library collection uploaded to Wikimedia Commons and shared on Wikipedia's sister projects. Based on the literature review and the analysis of the existing cultural content aggregators, its foundation grounds also in the experience acquired through the several research projects which have been con-

ducted over the last four years, GLAM visual tool and Wikipedia Primary School research project, in particular (see the appendix for more information about these projects).
The project investigates the ETH-Library collection which, with around 52.000 images uploaded from January 2016 to June 2018, is one of the biggest one ever uploaded to Wikimedia Commons. In fact, ETH-Library is one of the leading cultural institutions in the world in terms of contribution. Based in Zurich, it is also the largest public scientific library in Switzerland. The library collects eight million analogue resources and around 500.000 digital resources. The collection mainly covers the fields of architecture, engineering and natural sciences and it also includes historical photographs and topographic maps.
ETH-Library shares the collection through several interconnected channels, including ETH e-pics[1] - a digital image archive (containing also images protected by copyright) - and ETHorama[2] - a digital archive with geolocated images. Among the tools used by ETH-Library to monitor the spread of the collection over Wikimedia projects, there are Cassandra (only available for Swiss GLAMs[3]) and BaGLAMa2.
The collection is released under free licenses and the upload to Commons consists of aerial photographs by Swiss aviators coming from European, African and Asian cities. Map the GLAM tries to answer the following questions:

- What are the characteristics of images used within Wikipedia articles?
- Which are the main channels of the spread of images?
- What metrics can be used to measure the impact of a digitised collection?

The assumption is that by giving to stakeholders a set of visual tools to monitor the status[4] and the spread of the collection, it is possible to foster its usage (Cangiano et al., 2017).
Due to the big size of the collection, Map the GLAM has requested the implementation of several scripts to collect and process data coming both from the Wikipedia API or scraped via web pages. Scripts have also been used to generate the final data visualisations.
Map the GLAM is documented and released in open source on GitHub.

GLAM Culture Hub is an interactive prototype of a cultur-

al content aggregator based on the analysis of existing European cultural content aggregators and the results of the analysis of the ETH-Library collection on Wikimedia Commons and Wikipedia. The project tries to integrate several access points to the digitised collections in a simple and coherent user interface.
GLAM Culture Hub tries to answer the following questions: is it possible to apply existing guidelines for digital archives to cultural content aggregators? What are the interface characteristics that may foster the access and usage of digitised collections? I expect that the introduction of a visual tool as the main access point to the collection could be considered particularly useful by end-users.
GLAM Culture Hub is made with Invision - a prototyping web application which allows to add sensible areas on static images (see fig 4.18). The static screens of the project are also published on Wikimedia Meta.

4.1 Analysis of a digitised collection

The visualisation of a cultural collection aims to identify the dynamics that govern the access and use of collections within cultural content aggregators. Understanding the use of digitised content is difficult. Researchers applies several methods including both quantitative measures (bibliometric analysis and content analysis) and qualitative measures (stakeholder interviews, focus groups and questionnaires) (Meyer, 2009, p. 6). Despite big investment into the digitisation of cultural collections, and pressure from funding institutions to demonstrate the impact of the resources (Hughes, 2011, p. 2) it is often difficult to gain statistics regarding the use of the digitised collections (Terras, 2015). The comprehension of the impact, in fact, is a time-consuming activity and generally it can not be afforded in short-term projects (Hughes, 2011, p. 9).
Through the Map the GLAM project I tried to identify some metrics of the digitised collection within a cultural content aggregator and to provide access to stakeholders through the use of data visualisations.
The idea behind the project is to visualise data already available to understand the spread of an open collection. Among cultural content aggregators Wikimedia Commons - the multimedia repository of Wikipedia - is the most relevant for the purpose of the analysis because it is used worldwide and it

provides open data about the access and use of its content.
Map the GLAM is partially based on the results achieved through the GLAM visual tool research project (see "Related projects" on the appendix), the analysis of the most widely used GLAM tools (such as Baglama and GLAMorous), and an in-depth analysis of the most relevant cultural content aggregators in Europe. The existing GLAM tools are mainly based on the visualisation of page views. Despite a significant effort, from a technical point of view, these tools provide few details about the collection and its usage and their user experience is weak.
The project consists of three main steps: data gathering, data visualisation and validation with stakeholders. Before the project started, a map of the GLAMs contributing to Wikimedia Commons was made to select a specific cultural institution to be investigated. Among them, the choice has fallen on the ETH-Library[5] of Zurich because, with more than 52.000 images uploaded between January 2016 and June 2018, it is one of the leading cultural institutions in terms of files uploaded to Wikimedia Commons. Furthermore, its photographic collections have been used in several projects and linguistic versions.
Map the GLAM is published on a page within Wikimedia Meta[6] and the source code is also released in open source on GitHub[7]. Everyone is allowed to download and use the source code, with GPL-3.0 license, and the data, with CC0 license. The project has contributed to the identify of a set of people to be contacted for the end-user's survey.

Context
Wikimedia Commons is one of the projects funded by the Wikimedia Foundation. Among the other relevant projects, which are interconnected, there are Wikipedia and Wikidata (see fig. 4.1). Wikipedia is the largest and most popular web-based multilingual encyclopaedia edited by millions of volunteers.
Wikidata is a free and open knowledge base, an information system based on Linked Open Data (LOD)[8], structured and interlinked data which allow the users to make semantic queries. For the aim of the thesis, we only focus on Wikimedia Commons and on Wikipedia.
Wikipedia does not have strict rules, but it is based on five "pillars" which ground on free content and the adoption of a neutral point of view[9]. The 40 million Wikipedia articles, available in more than 300 languages, receive around 400

million unique visitors every month, with lower traffic during weekends and holidays[10]. Around 49% of the visits regards Wikipedia articles, and around 47% concerns images and other multimedia files (Reinoso et al., 2012).
Every minute, Wikipedia users make 120 edits, with a rate of two edits per second. Currently, there are approximately 4.000 active users, who make at least one edit every month. Individual contributors are called "Wikipedians". The majority of the Wikipedia community is composed by men (77%) (Herring, 2011). The policy states that Wikipedia is a secondary source, meaning that every information shown must be previously published elsewhere, such as in books, journals or online websites.
Wikimedia Commons and Wikipedia share the same typologies of pages. The most relevant are the article, file, category and template pages. The article page consists of content, references, multimedia files concerning a specific topic and it is the core of the encyclopaedia. It also lists a set of categories related to the topic. The user page is the profile page of the user, which is usually, a way to communicate personal information and intentions. The category page lists all the articles belonging to a category and it plays an essential role in keeping track of the articles related to a domain. Users add categories according to the content of articles: although the webpage shows categories as a tree, they are a set of tags. The template page is a preformatted content that can be filled and reused over multiple pages, such as the ones

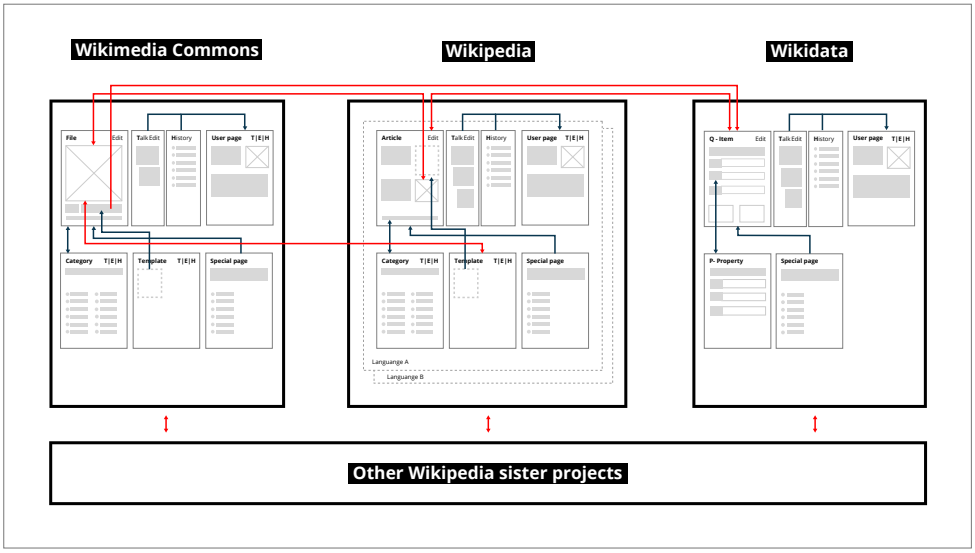


Fig. 4.1
Wikipedia information system.

containing information about people, locations and movies.

To quantify the fast growth of Wikipedia content and its fruition, Wikimedia foundation - the no-profit organisation founding Wikipedia and its related sister projects - defined several metrics. Wikimedia metrics are quantitative and qualitative statistics, such as the number of pages, visits and links, which are tracked manually or by computational tools. The most relevant quantitative metrics are related to the number of issues, incoming and outgoing links, page views, edits (in terms of added or deleted bytes) and editors.
The qualitative metrics concerns the quality and relevance of the article, which Wikipedia assesses using a grading schema.
According to the quality schema, the article class can distinguish articles ranging from a stub, an article that is developing but still incomplete, to a featured article which can be featured in homepage after an in-depth examination made by impartial reviewers. According to the importance schema, an article can range from low, if the subject is conceived for an audience with a specialist interest, to high, when the subject is a must-have for an encyclopaedia.
Due to the broad audience, GLAMs have a growing interest in cooperating with Wikipedia and its sister project to make their digitised collections open and largely accessible through the creation or editing of existing articles.

Data gathering and visualisation
The project mainly collects data through the Wikimedia Commons and Wikipedia API. A scraper and parser are built to gather and elaborate the data for the final visualisation. They both are Python and use the Panda.py and Num.py libraries[11]. The datasets and the code have been edited by using two code editors: BBEdit and Sublime text. The data are stored in JSON and TSV file formats.
Since Wikipedia changes very quickly, it is necessary to scrape data several times, in order to get them as updated as possible. Before starting an iteration, a script is developed and ran to check if every single image still exists.
The data gathering process follows these main stages: gathering the list of images, gathering the images metadata and gathering data about Wikipedia pages.
The list of images contributed by the ETH-Library is gathered by collecting images belonging to the category “Media contributed by ETH-Bibliothek”, because the ETH-Library

uses a unique account to upload and edit its images. In a correct protocol, following the Wikimedia Foundation specifications, the list of all the images should be extracted from the user uploads, with the aim is to collect data regarding the whole process of a digitised cultural object (from the digitisation to the online publication). Each image is provided with a variety of metadata, such as title, creator/artist, place, category, date, description, size (of the cultural object and of the digital surrogate), location/provenance, Exif data.

The data visualisation focuses on the status and the spread of the cultural collection and it mainly investigates the following aspects: the metadata of the images, the reuse in different pages and the gained visibility. Every visualisation encloses information about the dataset and the protocol - a set of actions to gather and visualise data. I decided to not rely on the data related to the use of categories. This is due to the fact that they are generally not used as a tool for accessing content but just as a way to get some contextual information[12].
The visualisations use a palette of colour according to the typology of elements displayed. In particular, there are files, files metadata, page typology, linguistic version, user typology and page views.
The data visualisation is an iterative process, as well as the data gathering. It consists of fast visual explorations using very basic visual models together with a diverging approach, and the implementation of specific visuals, which are based on the use of computational languages and a converging approach. During the visual exploration, RawGraphs[13] has been used as a tool to validate the hypothesis behind the investigated aspect of the collection. The final visualisations are performed by using the JavaScript library D3.js.
A set of visualisation protocols describes the steps to follow from data gathering to data visualisation, in order to make the data visualisations easy to replicate.
Map the GLAM also consists of a web prototype that aims to test a basic visual filter to access the ETH-Library collection. The visualisation shows the temporal distribution of ETH images and their related authors, subjects and mediums. It consists of a timeline and blocks showing the files of the collections. The data about the date, author and medium have been extracted from the Wikimedia user interface, while the data about the subject are obtained from the title and the description, through a semi-automatic process.

The user can explore the collection by clicking on the bars displayed in the timeline and on the blocks. A sidebar shows the list of the selected items and provides access to the digital resources in Wikimedia Commons. Due to the huge differences between the number of cultural objects over the decades, the vertical axis of the timeline uses a logarithmic scale.

Map of the GLAMs

I gathered the data on the existing digitised cultural collections on Wikimedia Commons in order to select the most relevant collection to be analysed (see fig. 4.2). I also geographically mapped the GLAMs to understand better which areas care more about open licenses. The steps conducted to make a map of the GLAMs contributing to Wikimedia Commons are the following.

1. Gathering the list of the GLAMs and the related number of files uploaded, starting from the following categories on Wikimedia Commons with 3 degrees of nesting:
    a. Files from content partnerships;
    b. GLAM Project;
    c. Commons partnerships;
    d. Files from batch uploading;
    e. GWToolset Batch Upload;
2. GLAMs with less than 50 files uploaded have been excluded;
3. Gathering the coordinates of the cultural institutions from OpenStreetMap[14];
4. Visualising the map by using D3.js and Leaflet, with tiles by OpenStreetMap;

The following is the top ten of the GLAMs contributing to Wikimedia Commons worldwide[15]:

1. Rijksdienst voor het Cultureel Erfgoed (heritage organisation);
2. Nationaal Archief (archive);
3. Library of Congress;
4. Naturalis Biodiversity Center (museum);
5. National Archives and Records Administration (archive);
6. German Federal Archive;
7. Metropolitan Museum of Art;
8. Deutsche Fotothek (library);

Fig. 4.2
The map shows the libraries contributing to Wikimedia Commons. The map includes 168 GLAMs: 3 are galleries, 32 are libraries, 26 are archives, 50 are museums, and 57 are other cultural institutions (foundations, public agencies etc.).

9. ETH-Library;
10. Tropenmuseum (museum);

Most of the GLAMs contributing to Wikimedia Commons are based in Europe. There not GLAMs releasing at least 50 images based in Africa. The typology of GLAM who release the most images is the library.

Images upload
The visualisation of the ETH-Library uploads aims to show uploads patterns and the adopted open licenses (see fig. 4.3). Here are the steps followed to collect and visualise the data concerning the pictures uploaded by the ETH-Library to Wikimedia Commons overtime:

1. Gathering the list of all the images uploaded by ETH-Library (via Wikimedia Commons API);
2. Gathering the upload date and the license of the photos (via Wikimedia Commons API);
3. Visualising the timeline by using D3.js.

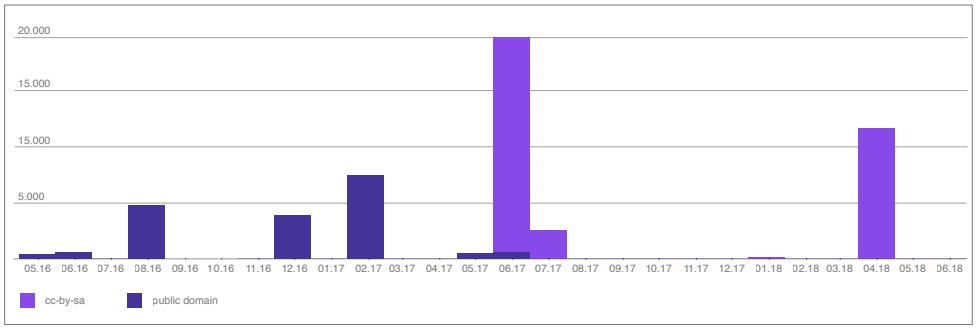


Fig. 4.3
The timeline (from May 2016 to June 2018) shows the uploads of ETH-Library images to Wikimedia Commons.

The visualization clearly shows the shift, in July 2017, in the use of license, from public domain to cc-by-sa. It also shows the huge release of digitised images.

Images size

The scatterplot of the digital and original images size aims to show the effect of the digitisation process in terms of preservation/alteration of proportions and dimensions (see fig. 4.4).
The steps conducted to gather and visualise data about the size of images uploaded by ETH-Library to Wikimedia Commons are the following:

1. Gathering the digital size of the pictures (via Wikimedia Commons API) starting from the list of all the images uploaded by ETH-Library (see Images' upload);
2. Gathering the physical image size by scraping the Wikimedia Commons page;
3. Parsing the data about the physical image size in order to make them easy to compare with the digital size;
4. Visualising the digital size and the physical size through a scatterplot, by using D3.js. The areas of the bubbles represent the number of items sharing the same dimensions.

The analogic scatterplot shows the dimension in pixels and it has many dots because of human crops images. The digital scatterplot, instead, shows the dimension in centimetres. The two main directions refer to portraits and landscape images. In the scatterplot on the right, there is only one main direction due to human errors in the entry of the real size. The big amount of bubbles in the digital size scatterplot is due to the fact that the images were cropped manually.

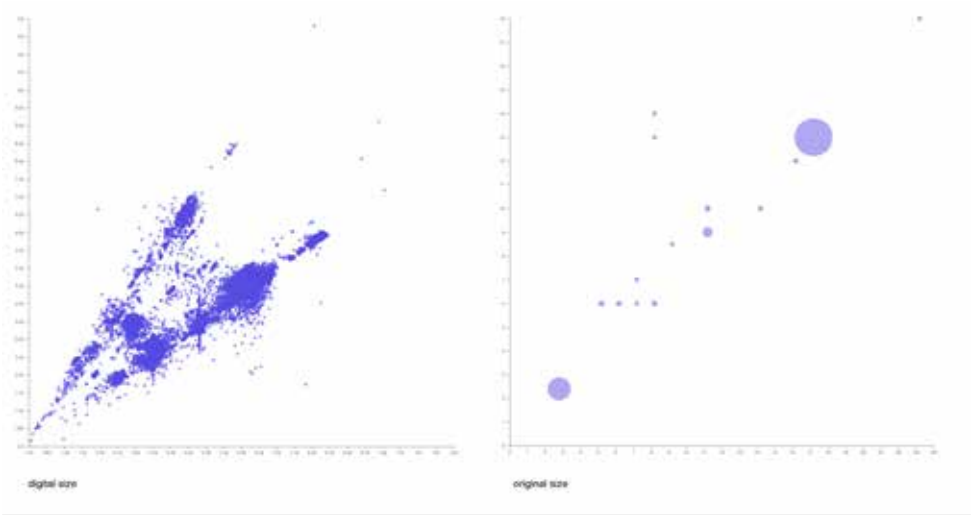


Fig. 4.4
The visualisation shows the digital and the physical size of the ETH-Library images. From the comparison, it looks like there is sometimes a swap of width and height in the physical size of the images (this might be due to the unstructured field "Dimension" in the images' metadata).

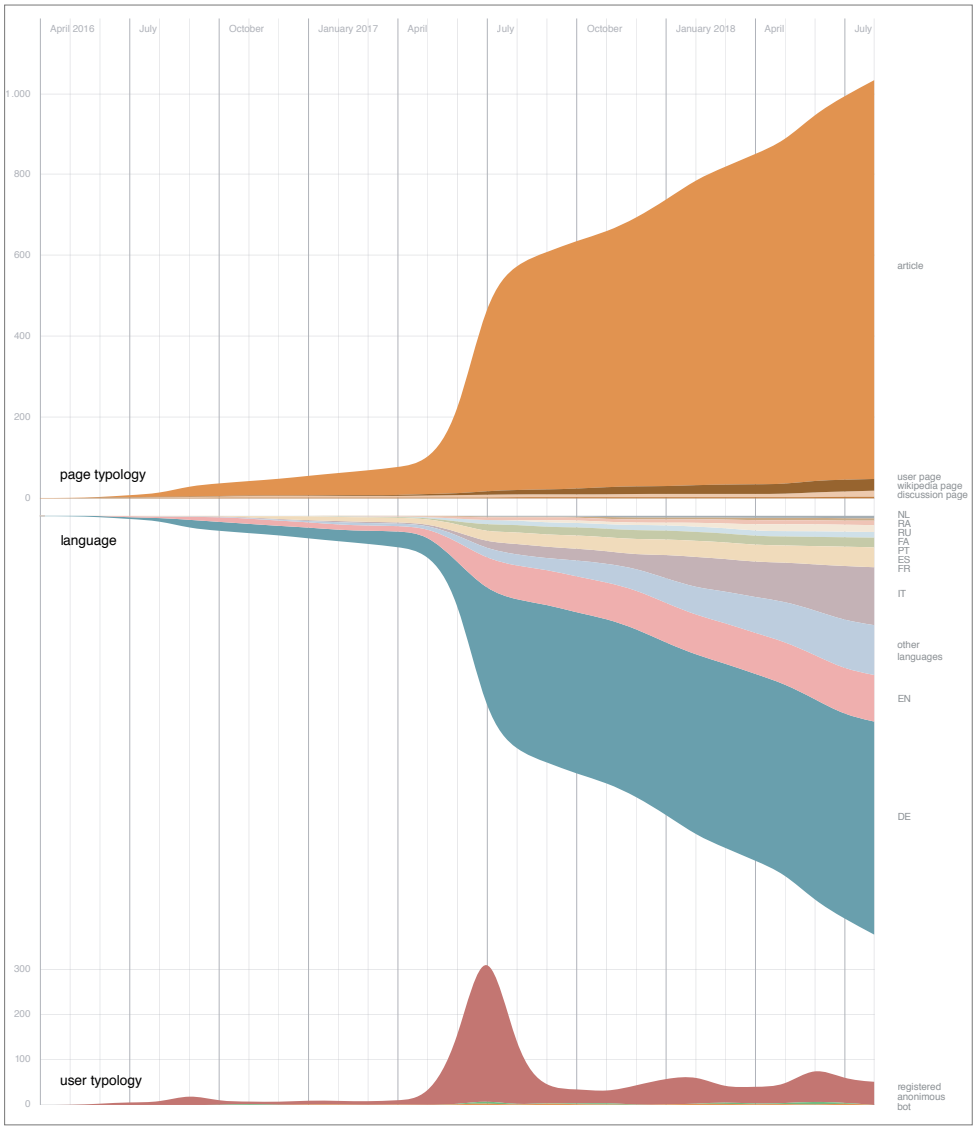
Images usage

The chart on the image's usage tries to show which the channels of the spread of the images are. This is a particularly relevant metric both for the Wikimedia Foundation and GLAMs (see fig. 4.5 and 4.6).

The following steps have been followed to collect and visualise data concerning images usage over Wikipedia and Wikimedia Commons:

1. Gathering the list of pages including the images by scraping the Wikimedia Commons page, starting from the list of all the images uploaded by ETH-Library (see Images' upload);
2. Collecting all the revisions of the pages, including the ETH-Library images via Wikimedia Commons API and Wikipedia API;
3. Parsing all the reviews to check the date when the ETH-Library image has been included in the page;
4. Visualising the Wikipedia and Wikimedia Commons timelines by using D3.js.

On the bottom, the charts feature the number of users who added ETH-Libraries images on different page typologies. The term "anonymous users" refers to users who made the edit without doing the login to Wikipedia. In this case, Wikipedia saves the edit under the public IP address[16] of the user. The term "registered users" refers to user who logged in to Wikipedia. Bots scripts are developed by the Wikipedia



community in order to perform repetitive actions over the platform. In the case of ETH-Libraries images, bots run to upload them on user and category pages.

Fig. 4.5
The visualisation shows the typologies of pages and their linguistic version in Wikipedia, where pictures were added.

Features of used images

The chart tries to identify which are the main attributes which facilitate the use of the image (see fig. 4.7). Here below the steps conducted to gather and visualise data about the images used in Wikipedia and Wikimedia Commons:

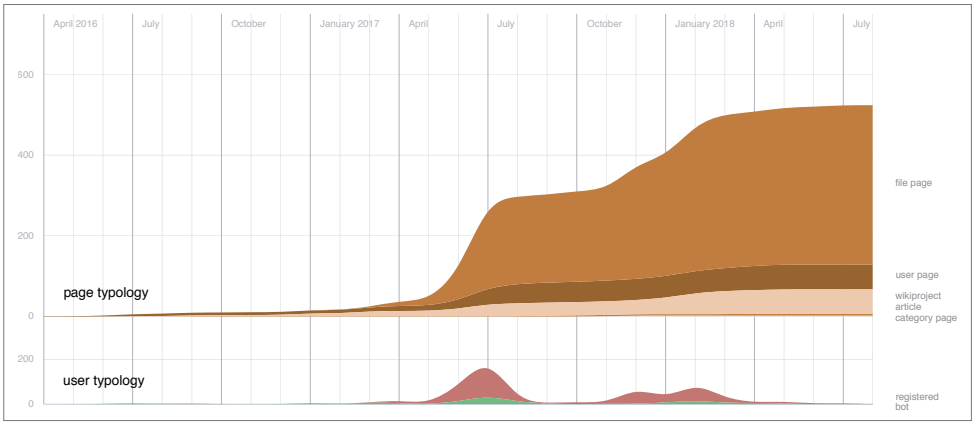


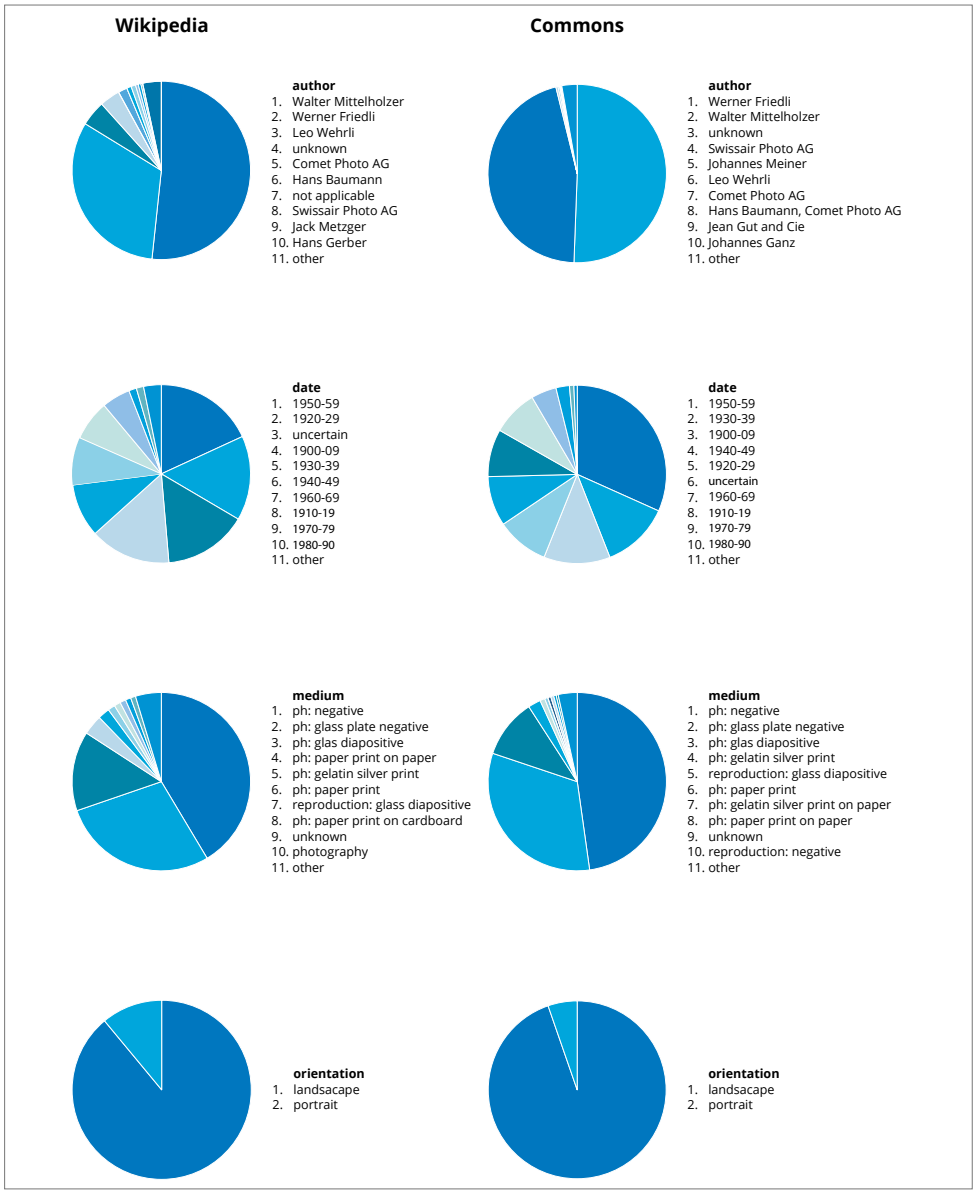
Fig. 4.6
The visualisation shows the typologies of pages and the related user typology in Wikimedia Commons, where pictures were added.

1. Resuming all the revisions of the pages including the ETH-Library pictures (see Images' usage) and the metadata of the images (see Images' size);
2. Parsing the data to obtain the list of images used in Wikipedia and Wikimedia Commons;
3. Parsing of the data to collect sets of authors, dates, mediums and orientation;

The ETH-Library images are mostly shared within Wikipedia Articles. The peak in the spread of images was due thanks to the contribution, during summer 2017, of a Wikipedian in residence who added images in German articles. Then, the spread keeps increasing thanks to the community who keep added the images in articles in German and other linguistic versions including English, Italian, French and Spanish.

The chart of page views is based on a research concerning the collection usage (Borowiecki and Navarrete, 2016), which aims to show the visualisation patterns for a better understanding of a possible correlation between the number of images per article and the page views (see fig. 4.8).
The following steps that have been followed to collect and visualise the daily page views of pages containing ETH-Library images.

1. Gather the daily page views via Wikimedia API, starting from the list of pages including ETH-Library images (see Images' usage);
2. Parse the data gained to obtain the average daily page views;



3. Visualising the chart by using D3.js.
4. Wikipedia and Commons mainly include overview of cities. However, Wikipedia has a large presence of portraits of intellectuals and other relevant people.

The visualisation shows that there is not a correlation between the number of images per article and page views. This

Fig. 4.7
The visualisation shows the metadata of the photos added to Wikipedia and Commons pages. Wikipedia pie charts refer to 5% of the collection. Commons pie chart refers to 50% of the collection. The numbers follow the clockwise order starting from the top.

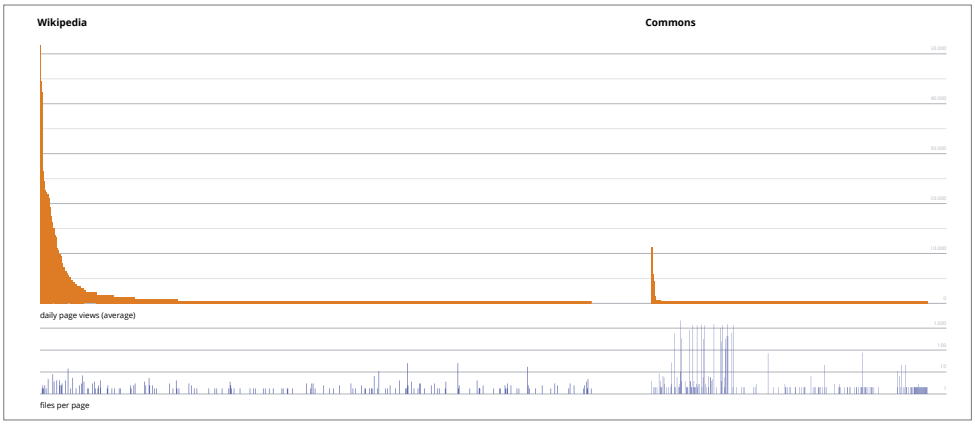


Fig. 4.8
The visualisation shows the daily page views of Wikipedia and Commons pages containing ETH-Library images. The average is calculated between July 2016 and July 2018.

means that images do not contribute to increase the popularity of an article and that the inclusion of an image in an article is not sufficient to increase its visibility.

Images position within pages

Following the previous chart, I analysed the average positions of images within different typologies of pages (see fig. 4.9). The aim of the chart is to understand how relevant ETH-Library images are for the Wikipedia Community. This parameter impacts the visibility of the images.
These are the steps conducted to gather and visualise the data about the images position within Wikipedia pages:

1. Gathering the HTML file of the page starting from the list of Wikipedia pages including ETH-Library images (see Images' usage);
2. Parsing the HTML file to obtain the starting tag of the ETH-Library image as a relative value, from 0 (the very beginning of the HTML page) to 100 (at the end of the page). In this step, data about the page typology (article/user page/discussion page/other types of pages) have also been collected;
3. Gathering of minimum, maximum and median values for every page typology.

This method does not collect data about the visual position of an image within a page but, after some manual checks, I found that it provides accurate results.

The chart shows that, on average, ETH-Library images are positioned within the first half of pages. This could mean

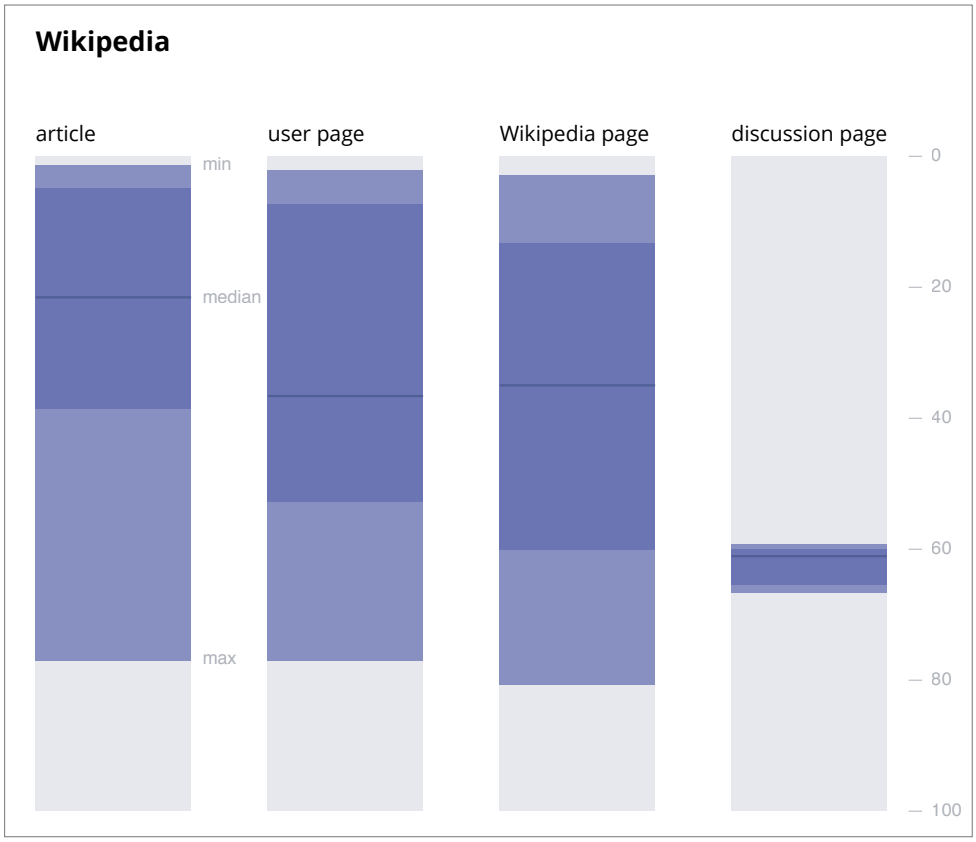


Fig. 4.9
The visualisation shows the average position of ETH-Library images within Wikipedia pages. The data are gathered through the analysis of the wikitext. It is not based on the visible position within the page).

that most of the end-users who access the Wikipedia pages containing the images actually see them.

4.1.1 Feedback gathering

The developing of the Map the GLAM project allowed me to gather data from the main stakeholders (the ETH-Library and the Wikimedia Foundation) and from the end-users who employed the ETH-Library images in their online publications. The feedback gathering essentially aimed to understand the metrics stakeholders might need to evaluate the online impact of their digitised collections and the issues end-users find while benefiting digital surrogates.
The feedback gathering was possible thanks to the publishing of the Map the GLAM project on Wikimedia Meta (see fig. 4.10), a wiki that collects information about projects related to the Wikimedia Foundation[17]. The page on Wikimedia Meta also allowed me to involve the community to

participate in the user research.
The visualisations produced have been discussed with several stakeholders on different occasions.
First, I gathered some feedback from some of my colleagues and friends working in the field of cultural institutions and promotion of open content. Then, during Wikipedia workshops and public events, I met people working for Wikimedia CH and Wikimedia IT. I also received a feedback from the responsible of the digital archive of the ETH-Library of Zurich. For him, Map the GLAM was an interesting opportunity to learn about the collection from another point of view.
During my period as a visiting student at the Urban Complexity Lab, in the Fachhochschule of Potsdam, I had the opportunity to receive a feedback from the members of the lab, as well as from people working for Wikimedia DE and Wikimedia Foundation.
Finally, I received feedback from the Wikipedia Community, after realising an article about Map the GLAM project in the Wiki-GLAM newsletter - a channel connecting thousands of cultural institutions and Wikipedia volunteers.

4.2 Designing a cultural content aggregator

GLAM Culture Hub (GCH) is an interactive mockup of a cultural content aggregator. Its goal is to better define the end-user's requirements and to evaluate, through an online

Fig. 4.10
Map the GLAM project on Wikimedia Meta.

survey, the usefulness of interface characteristics that may foster the access the use of digital surrogates. GCH is realised with Invision[18] - a prototyping web application - starting from the static images of the user interface.
The interactive mockup has been designed after having gathered the results of the ETH-Library collection analysis, after the analysis of the existing European cultural content aggregators and the definition of a set of design guidelines derived from these analyses.
GLAM Culture Hub features content coming from both Wikimedia Commons, in particular, the ETH-Library collection, and Europeana (see fig. 4.11). It aims to provide users with a mockup with features that may be implemented in a real cultural content aggregator for the access, navigation and use of digital surrogates. The design of GLAM Culture Hub is informed by the end-user survey and case studies analysis I conducted. During those analyses the following issues have been reported by end-users:

- insufficient or lack of information (especially about the terms of use);
- inaccurate information;
- missing of the digital surrogate (even though it has a description page);
- unclear or misleading classification.

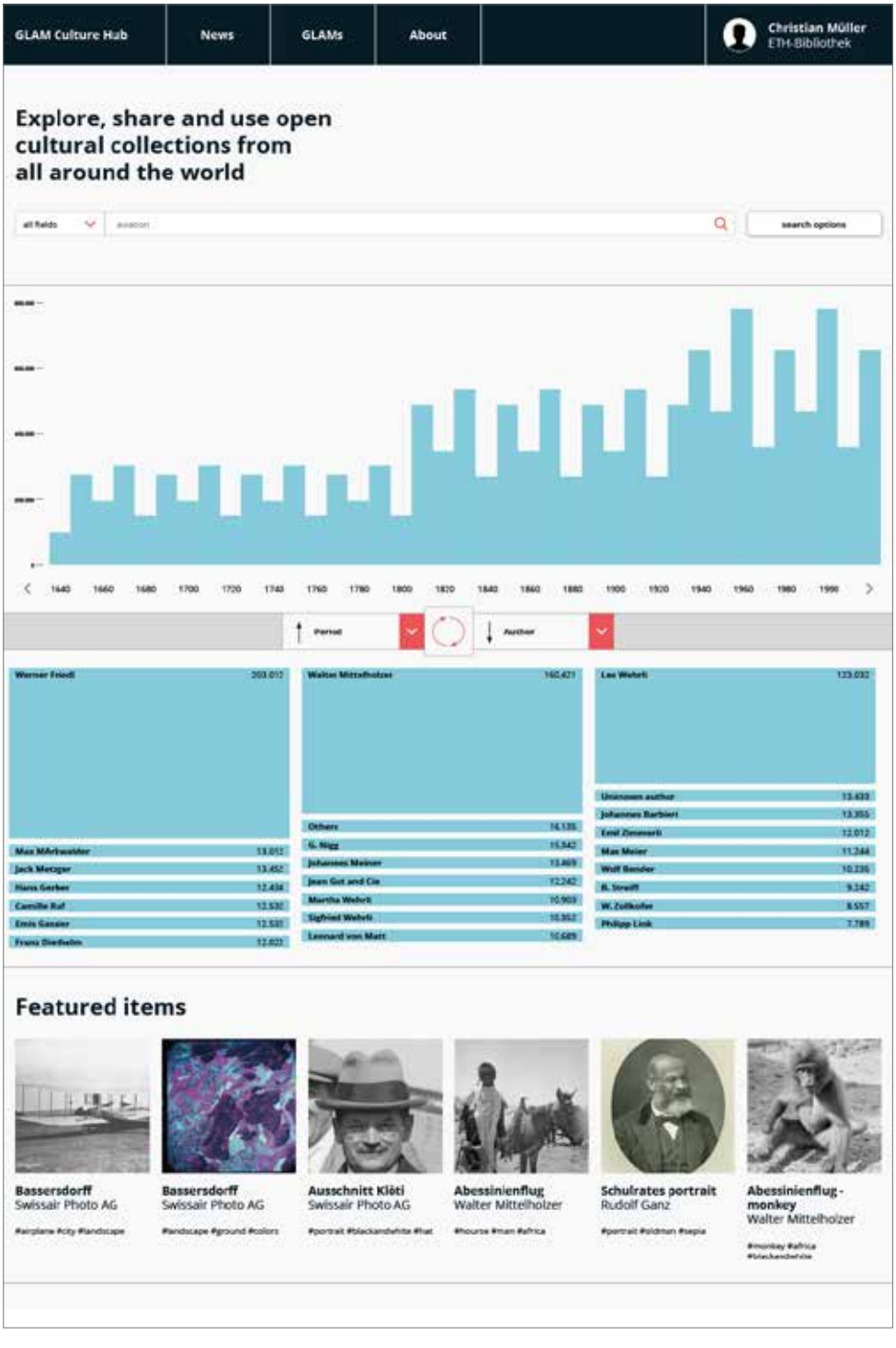
The current issues regarding the user interface are the following:

- limited search options
- inconsistent navigation between the main menu, lists and the screen with the single image.

Despite the issues, users have also reported that in recent years, content aggregators have enhanced their content and interface. GLAM Culture Hub uses the following design features in order to gather feedback about possible solutions on how to overcome the current usability issues.

- use of links among all the pages and to related surrogates.
- use of four options to narrow the search (period, GLAM, typology and author).
- use of a facet classification mainly based on descriptive metadata (not subject to possible interpretations by who categorize items)

Fig. 4.11
GLAM Culture Hub homepage.


GLAM Culture Hub
News
GLAMs
About
Christian Müller
ETH-Bibliothek
Explore, share and use open cultural collections from all around the world
all fields
search options
1640
1660
1680
1700
1720
1740
1760
1780
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
1999
Period
Author
Werner Friedl
203.012
Walter Mittelholzer
160.421
Leo Wehrli
123.032
Max Mörlswalder
13.012
Jack Metzger
13.452
Hans Gerber
12.434
Camille Ruf
12.532
Emin Ganzer
12.531
Franz Diedhelm
12.622
Others
14.135
G. Nigg
13.542
Johannes Meiner
13.469
Jean Gut and Cie
12.242
Martha Wehrli
10.903
Sighied Wehrli
10.352
Leonard von Matt
10.689
Unknown author
13.433
Johannes Barbieri
13.355
Emil Zimmerli
12.012
Max Meier
11.244
Wolf Bender
10.235
B. Streiff
9.242
W. Zollkofer
8.557
Philipp Link
7.789
Featured items
Bassersdorff
Swissair Photo AG
#airplane #city #landscape
Bassersdorff
Swissair Photo AG
#landscape #ground #colors
Ausschnitt Klöti
Swissair Photo AG
#portrait #blackandwhite #hat
Abessinienflug
Walter Mittelholzer
#horse #man #africa
Schulrates portrait
Rudolf Ganz
#portrait #oldman #sepia
Abessinienflug - monkey
Walter Mittelholzer
#monkey #africa
#blackandwhite

- clear display of licenses of use, with related icons
- use of a panel to allow end-users to propose a change or an integration.

GLAM Culture Hub essentially introduces some effective interface features on the design of cultural content aggregators, which have already been adopted on other generic content aggregators, digital archives and web applications. Among them, the distribution of filters on a horizontal line, to provide more space to the surrogates (existing cultural content aggregators display filters in a column).

Information architecture

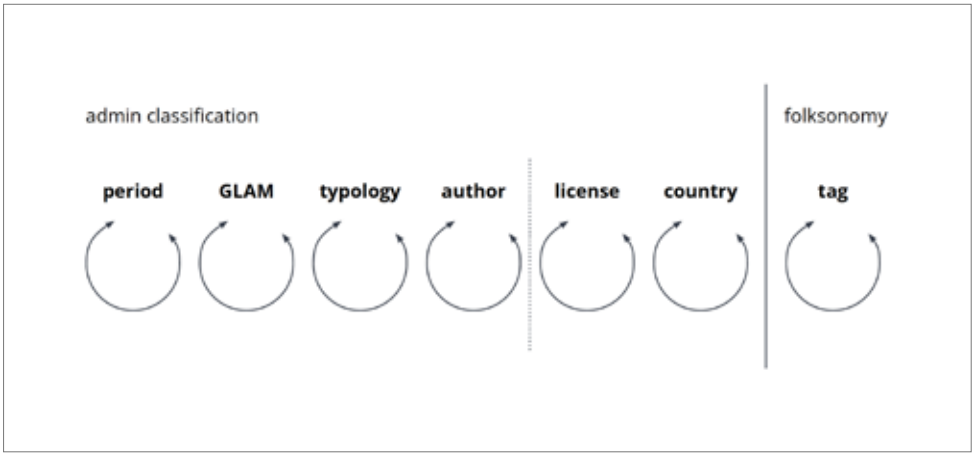
GLAM Culture Hub adopts a faceted classification system where GLAMs provide digital surrogates and related metadata and end-users can add tags (see fig. 4.12). The classification is based on the following metadata: GLAM, typology, author, license, location and tags.

GLAM Culture Hub consists of three main pages: the homepage, the list of items and the single item (see fig. 4.13). Among the other relevant pages, there are the GLAMs page (displaying them through a map and as a list), the user profile page and the GLAM page. Individual users can access the platform as unregistered users, registered users and as users working for a cultural institution.

User interface

The user interface of GLAM Culture Hub tries to integrate coherently three access modes: search, explore and browse. The search box is available on the top of every page. A chart

Fig. 4.12
Classification system schema. It uses a facet classification schema made by administrators and end-users (folksonomy).



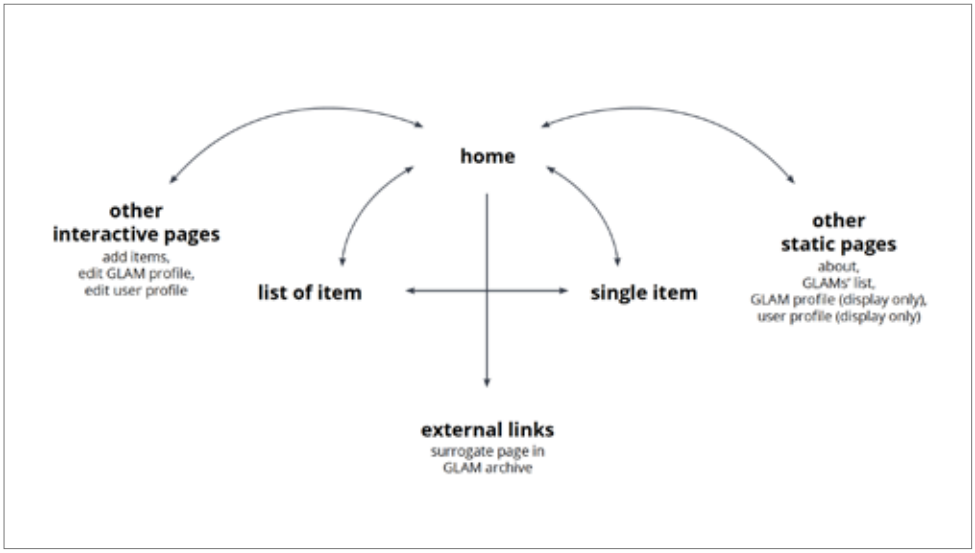


Fig. 4.13
Information architecture.

combining two visual models is applied to the entire catalogue on the homepage and to the individual collections in the GLAM pages. Interactive browsing elements are present on every page.

The search box allows users to search for a query in the whole set of available metadata or within a specific one. It also allows users to filter items by license and country. The chart consists of a timeline and stacker bars displaying respectively the number of items per date and the number of items per author. Two select menus allow users to switch between multiple data, such as typology, in order to facilitate the access to items for users who know different information about the artworks.

GLAM Culture Hub conceives the chart as a visual tool to filter items directly from the homepage. The interaction with the chart can lead to a single item or a list of items. The design of the chart is based on the guidelines regarding generous interfaces. GLAM Culture Hub displays a fake chart. However, a working prototype has been implemented and tested during the development of the Map the GLAM project: it displays real data coming from the ETH-Library collection on Wikimedia Commons (see fig. 4.14).

Interactive browsing elements mainly consists of sets of featured items, tags and GLAMs on the homepage. The single item page also features related items according to the author, location and date.

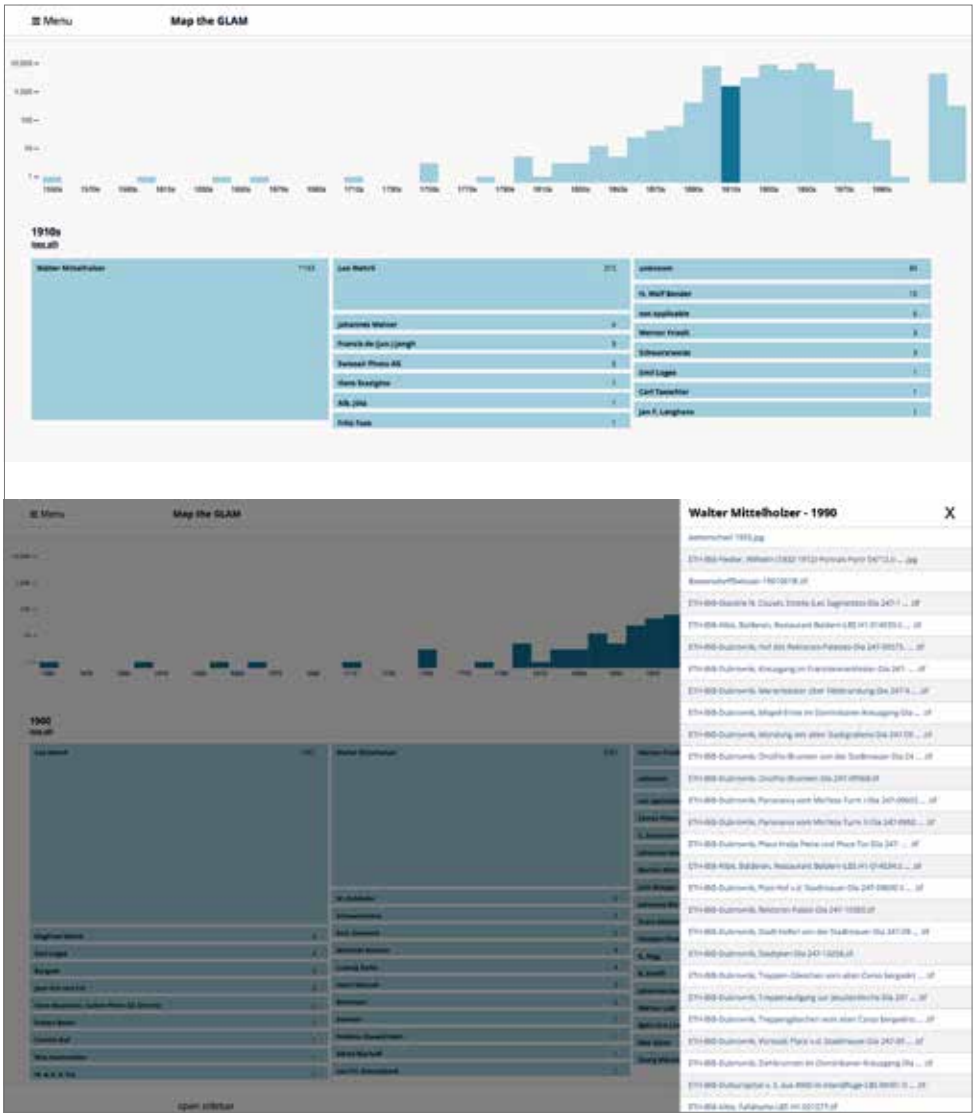


Fig. 4.14
Working prototype of the visual filter.

Within GLAM Culture Hub users can add tags to images (images that lack information are featured in the homepage to ask users contribution), create a personal collection and suggest an edit to the surrogates metadata.
From the visual point of view, GLAM Culture Hub has a layout based on a grid of 12 columns. The colour palette essentially consists of three colours: blue for the header and the footer, white for the background (so that it does not interfere with the images) and red for the interactive elements.

Every page includes interactive elements that allow users to access the collection and to navigate among surrogates (see fig 4.15). The following is the features list.

Homepage:
- Search form;
- Catalogue chart;
- Form to add tags.

Page with the search results (see fig 4.16):
- Search form;
- Sorting and display tools.

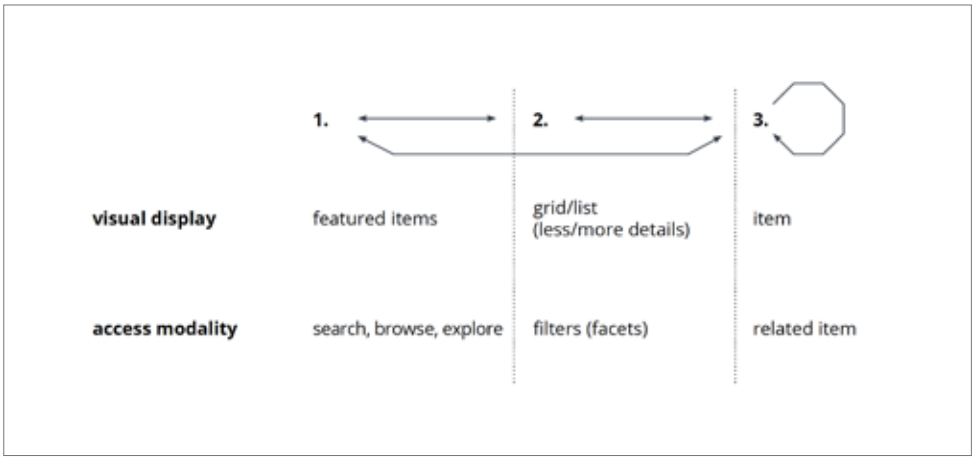
Single item page (see fig 4.17):
- Search form;
- Panel to add tags;
- Downloading and sharing tools;
- Form to suggest an edit;
- A tool to add a bookmark.

GLAM page (public):
- Collection chart.

GLAM page (private):
- Form to add items;
- Form to edit profile;
- Form to edit featured item;
- Catalogue chart.

User profile page:
- Information edit;
- Search form (applied to favourite items);
- Sorting and display tools;
- A tool to remove a bookmark.

Fig. 4.15
Access system schema.



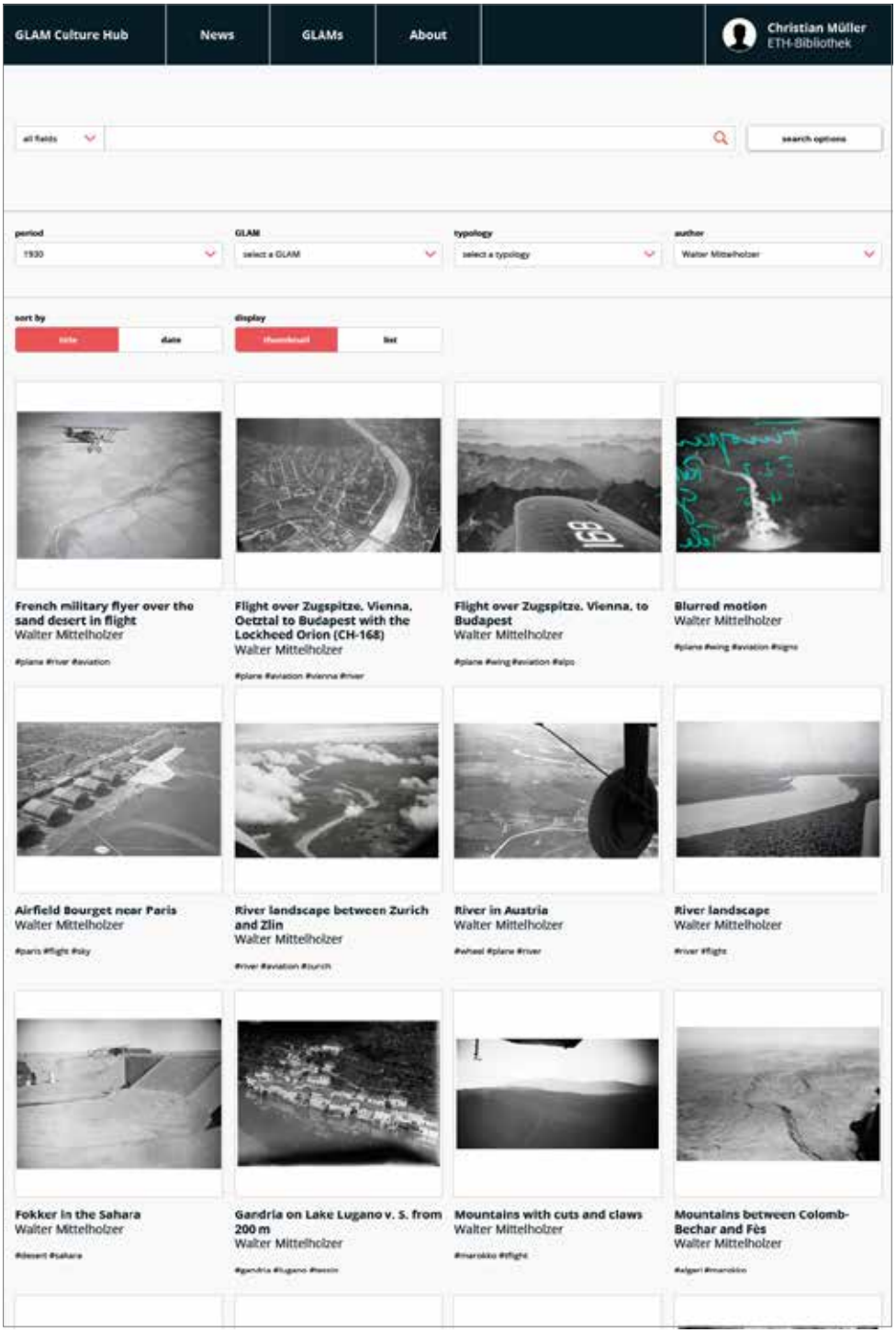


Fig. 4.16
GLAM Culture Hub, list of results.

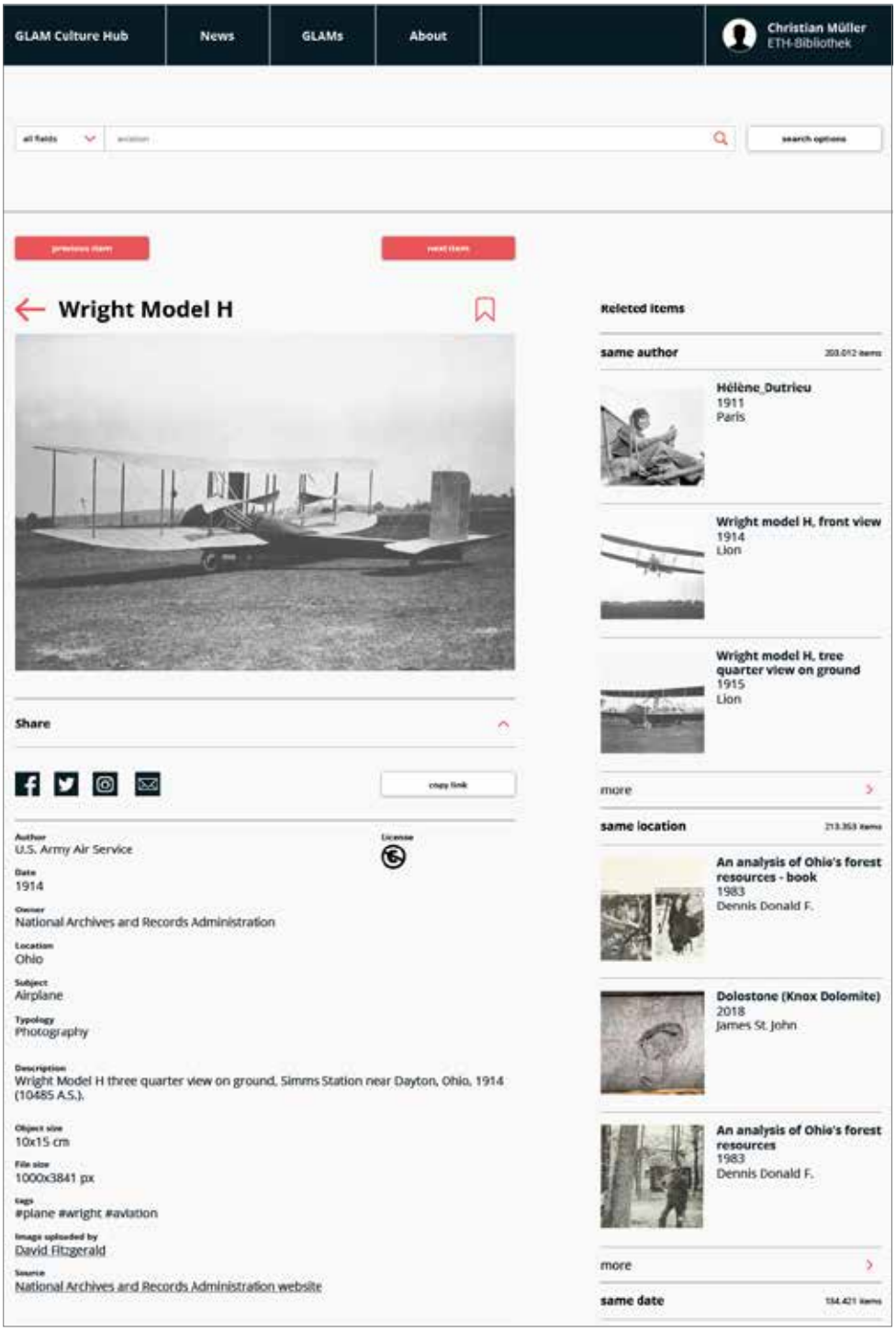


Fig. 4.17
GLAM Culture Hub, single item.

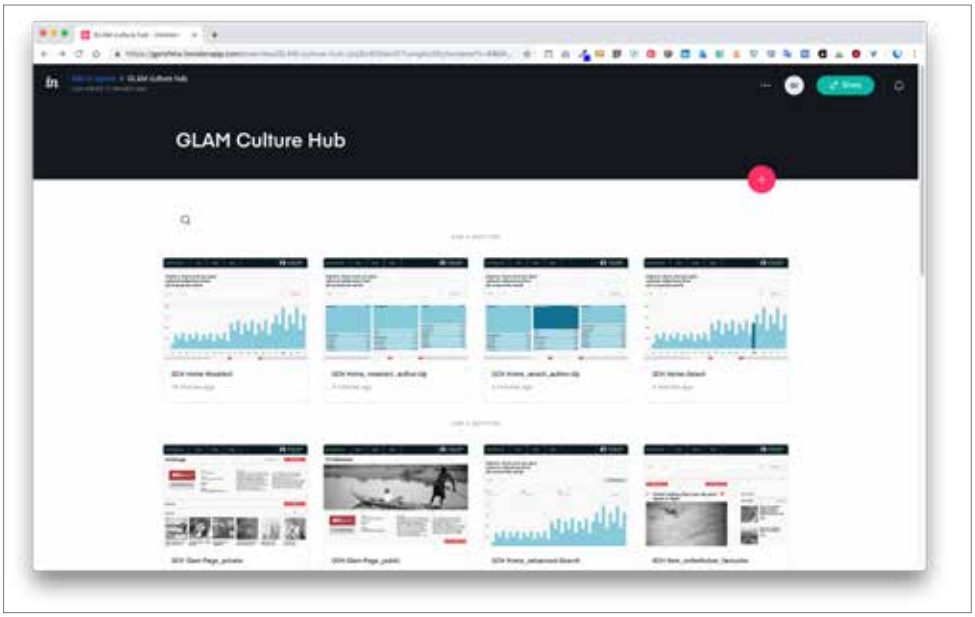


Fig. 4.18
GLAM Culture Hub on Invision.

4.2.1 Feedback gathering

Feedback on the interface of GLAM Culture Hub has been collected through an online survey, whichmainly adopts closed-ended questions. It also includes questions about user personal information, interface features and some other usability aspects. Questions concerning personal information mainly concentrate on the identification of the target, such as age, the higher degree and the frequency of usage of the open cultural content. Questions about interface features focus on specific characteristics of the cultural content aggregators, such as the search box, the chart and the GLAM map. Questions concerning usability aspects, instead, concentrate on the overall user experience and include open-ended questions, so that users can express their feelings and provide general feedback.
The survey has been advertised through several channels. Among them, there are my personal accounts on social networks (Twitter and LinkedIn, in particular), emails to people (identified through the analysis of the ETH-Library collection analysis) who often use cultural content aggregators and through an article on the GLAM-Wiki newsletter[19].

4.3 Results

The design project has two orders of results: the results related to the access of digital surrogates and the ones related to the usage. Regarding the usage of digitised collection, I tried to reply to the research questions.

What are the characteristics of images used within Wikipedia articles?
The most used images have a high resolution and are made by well-known authors. The subjects of the images are mainly portraits of relevant people, overviews of famous cities and other subjects well documented in the literature.

Which are the main channels of spread of images?
The main means to disseminate the digitized collections is the publication of articles on Wikipedia. Adding a relevant image in an high-quality article can give a great visibility to the image and its related GLAM. This is also due to the possibility for the community to generate multiple language versions of the same page.

What metrics can be used to measure the impact of a digitised collection?
From the experience acquired, it emerges that data on the use of images in multiple page typologies and languages can be useful for measuring the impact of a digitized collection (see fig. 4.6). The visualization of this data can be used for multiple purposes. GLAMs can understand the interest on the collection (in terms of language groups and user typologies) over time. The Wikimedia Foundation and the community may use the visualization to monitor the evolution of content related to the collection. The following paragraphs present further of results the two design projects.

4.3.1 Collection usage

The data visualisation, developed though the Map the GLAM project, has shown the vast extension of the ETH-Library collection regarding time, authors and mediums and a relatively low use within Wikimedia projects. Over the period 2016-2018, 8.800 users contributed to add around 1.000 images in more than 1.000 Wikipedia pages. The photos were initially added to the German version of Wikipedia articles, followed by the French, English and other 43 linguistic versions. The most used images portray intellectuals and oth-

er famous people, as well aerial views of European, Asian and African cities. The images are mainly added on pages about European and African cities. During the summer of 2017, there was a pick in the usage of pictures because there was a Wikipedian in residence at the ETH-Library. After a 400% increase in one month, the usage kept growing at a rate of 20% in the following months. As expected, the page views followed a long tail pattern, with an evident decrease in the page views during the summertime (June-September). In Wikimedia Commons, the head is much thicker and lower than in Wikipedia (very few pages are popular). The page views on Wikimedia Commons are much less relevant if compared with Wikipedia. Commons receive about 10% of the page views of Wikipedia, although it has 94% more of pages containing files.
Regarding the usage of the collection outside Wikipedia, ETH-Images are collected on image aggregators, such as alamy.com[20], gettyimages[21] and readtiger.com[22] and are mostly published on posts within history, travel and science blogs or articles.

4.3.2 Collection access

The GLAM Culture Hub survey has observed a general appreciation of the design by users.
Users expect to interact with a working prototype while, for time reasons, GLAM Culture Hub was made in fast prototyping. In many cases, I have intervened by sending further explanation via emails. In order to overcome the problem, I made a short video clip showing the main features of the project, and I shared it through my personal social network accounts and published on the Wikimedia Meta page of the project.
The questions that included textual responses were converted into a range of values from 0 (not useful - disagree) to 4 (extremely useful - strongly agree), to facilitate the analysis of the results.
People aged between 25 and 64, with a master degree, answered the survey: they are professionals working for cultural institutions (33%), designers (33%), photographers (11%) and other professionals (11%).
The search form is the interface tool that is considered the most useful (see fig. 4.19): it is considered extremely useful by 43% of respondents (with an average of 3.1 - very useful). Browsing tools are evaluated more useful than explorer tools,

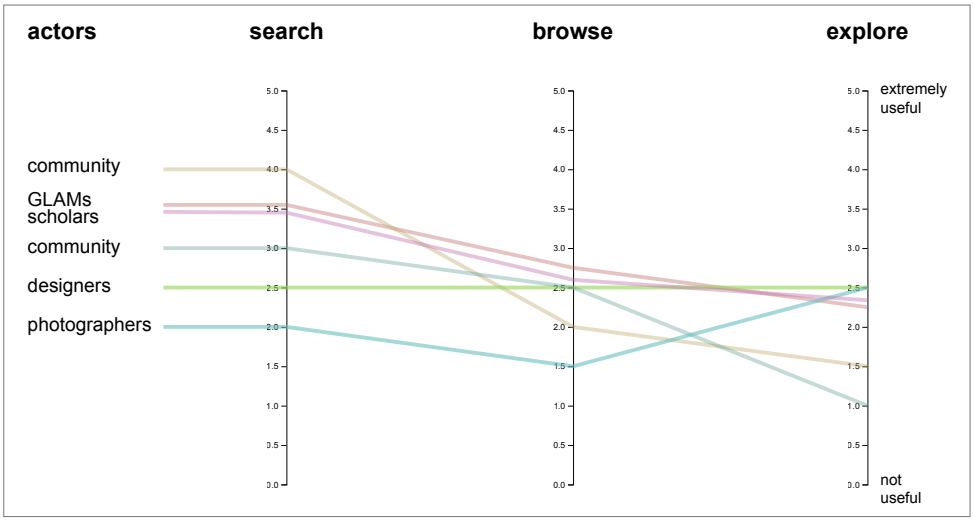


Fig. 4.19
Perceived usefulness of searching, browsing and exploring tools.

receiving an average of 2.4 (moderately useful), while explorer tools received an average of 2.3 (moderately useful).
Among the features that allow users to interact with content, the one that is considered as the most useful is the adoption of favourite items (with an average of 2.2 - moderately useful). The form to add tags to the picture and the form to suggest edits received the same rate in usefulness (2 - moderately useful).
Almost every respondent considers that information is clear and well organised on the screens. They also evaluate a consistent design throughout the system. Fifty percent of the respondents strongly agree to the fact that the various functions look well integrated in the system. Almost every respondent considers that the system is easy to use and would like to use it. Among the interactive elements, respondents like the downloading tools (with an average of 3.1 - like a little) and the charts (with an average of 2.9 - like a little) (see fig. 4.20).

4.4 Discussion

The analysis of the ETH-Library collection (containing around 52.000 items) has confirmed the previous research about cultural collections in Wikipedia. Wikimedia Commons uses 50% of the collection and Wikipedia only 5%. Twenty percent of images received 88% of views.
The project has led to the comprehension of the reasons why the usage is low in relative terms. First, many images

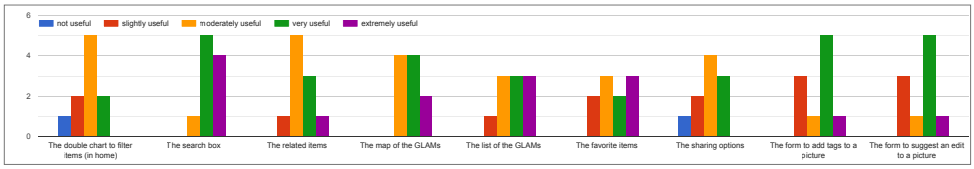


Fig. 4.20
Rate of the main interface aspects.

do not follow Wikimedia Commons policy since they are of low quality and there are no Wikipedia articles that would benefit from their inclusion. Secondly, most of the images is not documented in any online or paper publication. Thus, no articles could refer to them. In other words, there is a knowledge gap in Wikipedia, and there are no articles where images can be added. Finally, since Wikipedians search the images through the search box, only the pictures containing several and consistent metadata reach the query results. Browsing requires more time and the categorisation of content, sometimes, is misleading.

From the project, it emerges that the use of data visualisation may have a relevant role in mapping the spread of a digitised collection and providing GLAMs with insights concerning how to enhance the quality of their content. Data visualisation should comply with the following requirements: providing visual evidence on the visibility gained by cultural objects, showing the information gap and make the main channels of spread easy to understand.

The Map the GLAM design project showed some limitations related to the Wikimedia Commons information system. We can only refer to categories to track the digital surrogates belonging to a collection, but categories may change over time. Thus, in a long-term collection analysis, some other strategies should be found to keep track of the collection. Furthermore, the Commons system does not provide detailed data about page views.

Map the GLAM has shown the overtime usage of a digitised collection. In future works, the focus of the visualisation may analyse the individual usage patterns and the usage of cultural collections as a proxy to identify the knowledge gap in Wikimedia projects.

From the design project some issues related to the visualisations arised:

Conciliation of research interests

Wikipedians, Wikimedia Foundation and the digital curator have very different expectations from the visualisation of the spread of the collection. In particular, the Wikimedia

Foundation and the digital curator use evaluation metrics based respectively on the creation of content and on the impact they generate. While wikipedians and other end-users mostly rely on the visualisations as means to monitor the user activities (Profeta, 2018).

Fast evolution of Wikipedia
The rapid and constant changes within Wikipedia content and their classification make it difficult to obtain clean and structured data (Profeta, 2018).

Graphic consistency
Due to the large numbers of types of data available (including the metadata about the digitised images, the users who uploaded, shared and modified them and the articles where they are added), it is difficult to codify a coherent visual language (Profeta, 2018).

The evaluation of the GLAM Culture Hub prototype has shown a general appreciation for all the features that may allow users to contribute content to the cultural aggregator. These can be interpreted as an indicator of the fact that users love the idea to contribute to the web platform. The feedback collected reflects the requirements I had obtained through the stakeholders' analysis. In general, there is a need to provide end-users with features that not only allow the access to content but also its creation and management. From surveys, I expected to get more interest from users with regard to the graphical visualisation. Instead, data visualisation obtained controversial results. On one hand, users like it but, on the other hand, they do not consider it as the primary tool to access the catalogue. In my opinion, this is due to the fact that end-users have a clear idea of the surrogates they need. Thus, they prefer to use the search box to access the content quickly.
After receiving all the feedback, I realised that several elements would need improvement. Between them, the design of the chart, the list of results and the features to create a personal collection. I should also include a visual tool to see the online collection impact on the GLAM private page, and an intermediate value between "like a little" and "love it" in the survey.

Notes

1. ETH e-pics website: https://herb.e-pics.ethz.ch/.
2. ETHorama website: http://ethorama.library.ethz.ch/.
3. Cassandra website: https://stats.wikimedia.swiss/.
4. I use the term “status of the collection” because, within a cultural content aggregator, the collection can constantly be expanded and enhanced.
5. ETH-Library website: https://www.library.ethz.ch/en/.
6. Map the GLAM project on Wikimedia Meta: https://meta.wikimedia.org/wiki/Map_the_GLAM.
7. Map the GLAM on GitHub: https://github.com/giovannipro/map-the-glam.
8. Other relevant information systems based on Linked Open Data are DBpedia - a structured information system based on Wikipedia - and GeoNames - a system that allows users to add and retrieve geospatial semantic information to the Web.
9. The following link features the Wikipedia five pillars: https://en.wikipedia.org/wiki/Wikipedia:Five_pillars.
10. The following link features data about the linguistic versions of Wikipedia and their statistics: https://en.wikipedia.org/wiki/List_of_Wikipedias.
11. Panda.py and Num.py libraries are Python libraries for high-level data manipulation and analysis.
12. Information gathered from Ilario Valdelli, Community manager at Wikimedia Switzerland.
13. RawGraphs website: https://rawgraphs.io/.
14. OpenStreetMap website: https://www.openstreetmap.org.
15. The data about the GLAMs contributing to Wikimedia Commons refer to 2018.
16. The IP address is a numerical label assigned to each device connected to the Internet.
17. Map the GLAM project on Wikimedia Meta: https://meta.wikimedia.org/wiki/Map_the_GLAM.
18. GLAM CUlture Hub interactive mockups: https://gprofeta.invisionapp.com/public/share/UAWTM5SB8.
19. Link to the article about Map the GLAM on the GLAM-Wiki newsletter: https://outreach.wikimedia.org/wiki/GLAM/Newsletter/July_2019/Contents/Map_the_GLAM_report.
20. Alamy website: https://www.alamy.com/.
21. Gettyimages website: https://www.gettyimages.ch/.
22. Readtiger website: https://readtiger.com/wkp/en/.

Redactions, by Tim Sherratt, 2016.
Website: https://owebrowse.herokuapp.com/redactions/

"Redactions" is a project based on thousands of ASIO files of surveillance released under open license by the National Archives of Australia. The project focuses on the redactions made on the documents by an unknown archivist to keep some information secret. It consists of a digital archive which allows users to browse the documents through their redactions and to get some further detail on demand. The project is part of a series of projects based on the ASIO surveillance files. Part of the source files is available on GitHub.

5

Design guidelines

This chapter provides designers with guidelines for designing content aggregators complying with user needs. Guidelines are born from a literature review, user research and two design projects.

The implementation of design guidelines based on user research could have a positive impact on cultural institutions, communities and end-users. Cultural institutions invest large human and economic resources to digitise cultural objects. In the case of public institutions, resources come from citizens' taxes and few or no access to surrogates lead to a waste of resources. Communities contribute to the improvement and spread of the content. Their activities as volunteers, as well as the growth of the community, are closely linked to the interface tools for accessing and using the contents. The more these tools are simple and powerful, the best the community can perform its activities. Easy-to-use tools for accessing and using this content allow end-users to achieve their goal and they also increase the visibility of both the cultural content aggregator and the cultural institutions partners. Moreover, the definition of guidelines is a contribution to the research in design. In fact, in the literature, there are no other detailed guidelines targeting cultural content aggregators and the existing ones mainly focus on digital archives and the use of visual tools for exploring the collections (Whitelaw, 2012; Dörk et al., 2012).

The guidelines are a tool to support the design of cultural content aggregators. In particular, they aim to foster access and use of digital surrogates and avoid usability issues. The

guidelines are thought for designers. They are based on the needs of relevant end-users of cultural content aggregators (see paragraphs "Writers" and "Communication agencies" on chapter 3.2). Furthermore, since the design guidelines define end-users' requirements and propose solutions that can positively impact the visibility of their collections, they can also be useful for cultural institutions.
The guidelines do not offer details for the design of a particular interface tool, but at the highest level, they suggest design strategies be adopted to meet end-user needs. Although during the research the needs of cultural institutions were considered, for the design of visual tools for monitoring the use of collections, the design guidelines aim to satisfy end-user needs.
The guidelines knowledge coming from existing literature in cultural analytics and computer science and user research mainly focusing on a design project. They are mainly conceived for the design of cultural content aggregator, but they can also be used as a tool for driving a redesign.

The design guidelines are inspired by the OpenGLAM principles[1]. They consist of five principles that define the actions a cultural institution must perform to open its collections and are grounded on existing policies regarding cultural content.
The OpenGLAM principles were published in April 2013 by the Open Knowledge Foundation, within the framework of the OpenGLAM initiative[2]. OpenGLAM was co-funded by the European Commission and promotes free and open access to the digital cultural heritage held by Galleries, Libraries, Archives and Museums. OpenGLAM has several local working groups which promote the application of OpenGLAM principles. Among the more active working groups, there is OpenGLAM CH[3].
Here below the OpenGLAM principles.

1. Release digital information about the artefacts (metadata) into the public domain using an appropriate legal tool, such as the Creative Commons Zero Waiver. This promotes the maximum possible reuse of the data and allows your resources to become more discoverable.
2. Keep digital representations of works for which copyright has expired (public domain) in the public domain by not adding new rights to them. This promotes the maximum possible reuse of the content.

3. When publishing data, make an explicit and robust statement of your wishes and expectations with respect to the reuse and repurposing of the descriptions, the whole data collection and the subsets of the collection.
4. When publishing data, use open file formats, which are machine-readable. The structure and possible uses of the data should be well documented, for example in a datablog or webpage.
5. Opportunities to engage audiences in novel ways on the web should be pursued. Give opportunities toyour audience to curate and collect items from your collections. Where possible, consider allowing your users to enrich and improve your metadata by leveraging crowdsourcing applications.

The utility and adoption of the OpenGLAM principles have been analysed through a survey conducted in 2018 by the OpenGLAM initiative[4], which showed that principles are not well known among cultural institutions. A vast majority of institutions who knows the principles claims that they are useful for its work. Among those that did not consider these principles useful, most of the critiques concern the lack of support from cultural institutions. They also report a lack of guidance on how to apply the principles in practice.

Starting from the OpenGLAM Principles, this chapter proposes the OpenGLAM Design Guidelines. They aim to provide designers of cultural content aggregators with actionable guidelines that may foster the openness of digital surrogates. The guidelines are intended as complementary to the OpenGLAM Principles and the related GLAMs activities.
The following are the OpenGLAM Design Guidelines that will be further explained in the related chapters:

1. Adopt multiple classification modalities.
   a. use direct, indirect and computational classification methods;
   b. adopt facet classification systems based with a cluster schema;
2. Provide several access points
   a. provide end-users with search box with multiple search options;
   b. provide end-users with options to browse the collections;c. provide end-users with a graphical rep-

resentation of the collections;
3. Make connections among content.
   a. at a general level, create sets of surrogates with similar characteristics;
   b. at a detail level, create connections among a single surrogate and related items;
4. 4. Support the use of surrogates.
   a. provide end-users with high-quality images;
   b. provide end-users with detailed information about the characteristics and the provenance of the digital surrogate;
   c. provide end-users with tools for collecting and managing surrogates;
5. 5. Encourage user contribution.
   a. provide end-users with tools to generate and edit content (such as articles and collages)
   b. provide end-users with tools to classify digital surrogates
   c. provide end-users with tools to leave comments and feedback regarding content

5.1 Adopt multiple classification modalities

Classification is an essential operation to retrieve digital surrogates within a cultural content aggregator and it is based on the metadata of surrogates. In fact, the gathering of these data, allows the inclusion of single items within groups, or classes, that will constitute nomenclatures, taxonomies or facets (see chapter 3.3.1). The following example tries to clarify the relationship between metadata and classification.
We have a group of digital surrogates and each of them has only the name of the author as metadata, thus the only way we can classify surrogates is by author, in an ordered list (also called nomenclature). Every item of the list will contain the author's name and its related surrogates. If we have both author's name and the date of creation as metadata, we may classify surrogates in a taxonomy consisting of a hierarchy with the artistic period, style and author. Otherwise, if we have as metadata author, date and medium, we can classify surrogates in a more complex system with multiple facets. Thus, the more metadata we have, the more articulated the classification will be, allowing allowing to understand the aspects that characterize an item and its context. This point is particularly relevant because of the logic behind

cultural content aggregators, which are different from digital archives.
On one hand, digital archives generally adopt curatorial strategies, which consist of a specialist work of selection, organisation and display of digital surrogates following the interpretation of cultural objects. Curating a digital archive is a time-consuming task, performed by the digital curator, that may include exceptions in the classification and the display of digital surrogates. The person in charge of curating a digital archive is the digital curator.
On the other hand, cultural content aggregators adopt aggregation strategies based on an automatic clustering of digital surrogates, according to similar characteristics extracted by their metadata. In a cultural content aggregator, the classification and the representation of surrogates have no exceptions and algorithms perform the aggregation process in a short amount of time. Due to the characteristics of content aggregators, cultural institutions should provide digital surrogates with the whole set of metadata available, written accurately and following the aggregators standards.

Context
The classification consists in the creation of classes of items based on their metadata (see chapter 3.3.1). Classification modalities refer to the process of classification of collected items. Within cultural content aggregators, we can identify three classification modalities: direct, indirect and computational.
Direct classification refers to the classification of items without intermediation. It is made by the administrators, a few people that have an authoritative role within the information system, and can generate any type of classification scheme. Direct classification is potentially the most consistent classification modality, and usually it is based on a set of predefined metadata. Cultural institutions usually perform direct classification.
Indirect classification refers to the items' classification conducted through an iterative process. It is generally called folksonomy, and it is made by many people with the same capabilities within the information system through the use of categories or tags[5]. Indirect classification can be more detailed than the direct classification providing a broader context to the digital surrogate. General users who create an account on the cultural content aggregator perform in direct classification. It is adopted to generate faceted classification

systems (see paragraph “Facet (graph and cluster schema)” on chapter 3.3.1) and mostly uses descriptive metadata.
Computational classification refers to the automatic classification of items. Computational classification algorithms classify items according to predefined rules (in the case of scripts) or through dynamic learning processes (in the case of machine learning). Computational classification uses descriptive and structural metadata, and generate classifications with cluster schema. It is made automatically according to rules predefined by the aggregator’s administrators and It is incredibly detailed, and it is mostly based on surrogate visual properties.
From the research, it emerges that the users need a clear and simple organization of the contents, which correct display its classification scheme (see paragraph “Writers” on chapter 3.2 and chapter 4.3.1).

Design strategies
Using multiple classification modalities may allow digital surrogates to gain rich and consistent metadata. Since the research has led to the conclusion that cultural institutions end-users need a clear classification of surrogates, based on metadata produced by both human and computer (see chapter 3.2), the cultural content aggregator should use multiple classification modalities and a facet classification. Every classification modality should be based on a specific set of metadata (see tab. 5.1).
The classification systems should be based on accurate descriptive, structural and administrative metadata. Thus, cultural aggregators should use faceted classification systems and multiple classification modalities. Faceted classification is simple and highly flexible because it uses independent classes. The adoption of multiple classification modalities is compatible with the faceted classification and allow the generation of extensive metadata over the three typologies. The lack of a clear classification system and accurate metadata may lead to incomplete or misleading research results.

In the three classification modalities, errors may be erased. Thus, it is crucial to provide GLAMs and end-users with easy interface tools, in order to report and fix classification errors. Furthermore, users should be allowed to provide missing metadata or propose an edit.

	Descriptive metadata	Structural metadata	Administrative metadata
Direct classifica-tion (by adminis-trators)	Date Author Subject	Medium Size of the cultural ob-ject	The owner (cultural in-stitution name) Place License of use
Computational classification	Computer-generated tags	Size of the digital surrogate Colour palette	-
Indirect classification (by folks)	User-generated tags	-	-

Tab. 5.1
Table of the classification modalities and related metadata

User interface
In the use of multiple classification modalities to generate a facet classification, the user interface has two main goals: make the classification system easy to understand and ensure the maintenance of consistent metadata.
To make the classification system easy to understand, the user interface should comply with the following requirements. Firstly, the user interface should be able to distinguish the different metadata typologies (descriptive, administrative and structural metadata). Secondly, the user interface should visually clearly communicate the adopted classification system (see chapter 3.3.1). Finally, the user interface should always return results of a query or a visual filter. If a user search generates no surrogates, a message error should be displayed together with hints to continue the navigation. To maintain the consistency of metadata, forms for metadata entry should provide placeholders as an example of the type of content and its format, and suggestions based on existing data, and adopt some physical constraints to avoid errors. administrators should have the possibility to insert data in batch. While end-users should be allowed to insert or to propose a metadata once at a time to avoid malicious actions.

5.2 Provide several access points

Allowing easy access to the digitised collection is the primary goal of cultural content aggregators. They were born from the need to gather in one single place digital surro-

gates coming from multiple cultural institutions. The modalities to access digital surrogates within a cultural content aggregator should reflect the experience of visiting a classic library. Within a library, users may ask for a book to the librarian, consult the list of books (in alphabetical order of title or author), wander among the shelves waiting for any inspiration coming from the back of the books.
The library offers multiple access points to content and great flexibility in the procedures to reach an item. Probably, within a traditional library if users can access books only by asking the librarian or if they have to reach a shelf without any visual reference, they may give up to rent a book. Thus, the knowledge remains obscure, and the library fails its primary goal. In the access to surrogates, cultural content aggregators should be considered as augmented libraries meaning that it should provide end-users with multiple, quick and simple option to obtain relevant and punctual results.

Context
Current cultural content aggregators mainly provide users with the search box (the librarian in the classic library). Even though the search box is fast and powerful, it cannot fit any user and need. The users of content aggregators, in fact, may have different backgrounds and levels of expertise. Thus, they need different entry points to access collections.
As well as cultural institutions are generous in providing to the community a significant quantity of digital surrogates, as graphical interfaces should be generous in providing users with multiple ways to access these content (see chapter 3.3.2). The research has demonstrated that only one way to access collections is not sufficient. This is due to the fact that different end-users have different needs and expertise (see chapter 3.2). There are three modes to access content: search, browse and explore. The search mode always requires the user to operate actively in order to access a specific content, while in the browse and explore modes the user has to be active while randomly navigate the content. The user interface of a cultural content aggregator should allow the three access modes, because an interface that gives access to digital surrogates in few ways might restrict the content use. It also might not meet the users' needs and force them to leave the platform.

Design strategies

A cultural content aggregator must should allow end-users to access digitised collections through an integrated model consisting of search, browse and explore modes. Creating an integrated model to access the surrogates requires the design of a linear information architecture and easy-to-use access tools.

The information architecture should be based on three main page templates, which are interconnected with each other: the homepage, the list page and the surrogate page. The homepage provides access tools belonging to the three modes. The list page features the results according to the user's request, thus it should offer direct access tools in order to narrow the user search (search tools). Finally, the surrogate page presents a single item, thus it should provide indirect access tools to expand the search (explore and browse tools). Every access mode has its own tools providing more access points.

The search mode is based on the search box. Since the search box allows to explore for a query at a time, we could consider that it consists of one access point. It provides both simple and advanced functionalities. A simple search mode search for a query in the whole set of metadata available, while an advanced search box does it in a specific sub-set of metadata. For instance, in a simple search box, the query "Leonardo Da Vinci" will return both artworks of the Renaissance artist and his portraits. On the other hand, an advanced search box may return one of the two types of results or the results split into the two types. It may also provide an autocomplete function to reduce time and user's cognitive effort. Both simple and advanced search box should provide relevant results that can be filtered further.

The browse mode is based on interactive elements, which allow access to sets of content or specific surrogates. They are usually introduced by the texts "featured items or collections", "list of contributors", "list of disciplines", "browse by medium" and "similar items". Since the browse mode consists of an array of elements that may interest the user, we could consider that it provides multiple access points. In the literature, there are no codified rules about how to design browse tools. However, it is useful to have browsing tools providing a general overview of the cultural content in the homepage and tools showing specific, related items in

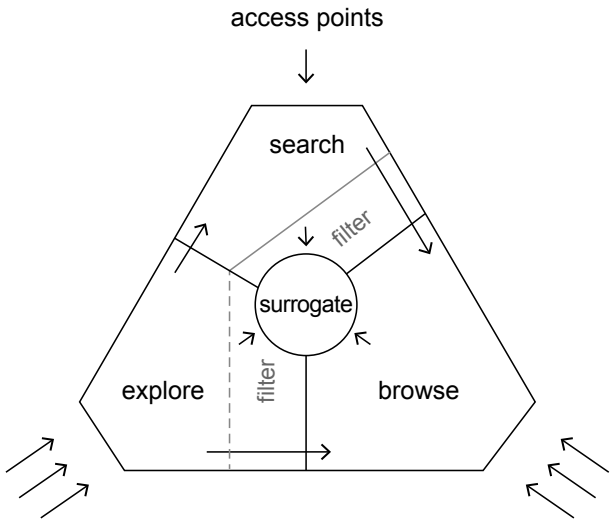


Fig. 5.1
Proposal for an integrated access model.

the surrogate page. Tools for browsing the content should give access to surrogates that receive few visits. Content aggregators should not feature only the most popular items, in order to enlarge the number of items seen and used.

The explore mode is based on visualisations of metadata the surrogates and allows users to reach the surrogates by a direct manipulation of the visualisation, providing multiple access points. The explore mode best suits an individual who have a vague idea of what to look for, because it focuses on surrogate's characteristics that the user might know. An explorer tool must be able to switch from a general overview of the collection to surrogate details in a fluid way.
Among the visual models for the representation of metadata the surrogates, the timeline is one of the most relevant because it gives access to the context, it is widely known and it fits with the requirements of the web browsers.

User interface
Within a cultural content aggregator, providing multiple access points, it is fundamental to integrate the access tools according to their characteristics in a clear visual hierarchy. Since search tools allow direct access to the content, they should be placed in a prominent position. Explorer tools, instead, can stay on a secondary level, but they may require a wide space. Tools for browsing the content can have a secondary role, as well as be spread over the user interface.

Both tools belonging to the search and explore mode may need textual or visual filters to narrow the search, which may be shared among the two modes.
In the proposed integrated model, tools belonging to different access modes can coexist within a unique access tool. The combination of search and browse mode is the most common. In this integrated tool, a user can make a query through a search box and then browse the results, while in a search-explore combination, there are two options. On one hand, search filters a data visualisation and, on the other hand, the visualisation filters the results coming from a search query. Finally, in an explore-browse combination, the visualisation integrates browsable surrogates.

5.3 Make connections among contents

In the cultural heritage, everything is connected. Every artwork mentions or is inspired by another one. In art, literature, fashion and other artistic sectors, cultural objects are born from the hybridization of multiple existing artworks.
In art history, there are several well-known examples of connections between current and previous artistic periods. For instance, the Renaissance popped up after the rediscovery of the classic period in the late Middle Ages. Romanticism was an artistic and literary movement originated from the ideas coming from the medieval period. Even avant-garde works, in the early 1900s, have a connection with previous artistic movements. For example, the artworks by Giorgio de Chirico were inspired by classical art, while Pablo Picasso, took inspiration by cultural artefacts of primitive people.
Human artworks and knowledge come from cultural sedimentation processes and confrontation among artworks produces new content and ideas.
Unfortunately, from the research, it emerges that European cultural content aggregators do not make the complex network of relationships among artworks visible. Individual artworks often lack the context of provenience and clear explanation about other related items. In most of the content aggregators, due to several restrictions affecting the searching and browsing tools, using related items in the surrogate page is the only way to access items sharing the same artistic period or location. This may generate an incomplete perception of the cultural heritage among general users and make the analysis difficult for scholars and other experts.

Context

Within cultural content aggregator, the term "connections" could be used to define textual or graphic elements that allow end-users to observe a relationship between two or more surrogate metadata or between two or more digital surrogates. Connections can be based on the three types of method: descriptive, structural and administrative. Connections based on descriptive metadata join surrogates with characteristics in common (such as the author and subject). Connections based on structural metadata connect surrogates with similar visual appearance and those based on administrative metadata connect surrogates with characteristics of the provenance in common (such as the place of origin). The visual elements that define connections can be both static or interactive.

The display of connections among digital surrogates has a relevant role for several reasons. It fosters the access to content, enhances the comprehension of the context of the provenance of artworks and facilitate the discovery of other related topics. Connections are particularly relevant for end-users who need interface tools to compare different items (see paragraphs "Writers" and "Communication agencies" on chapter 3.2). The general audience usually does not need the references that inspired the works of an artist. Thus, it is sufficient to provide them with items and contextual information.

The literature states that interfaces should provide end-users with interactive visual elements that allow them to move among sets of surrogates in a simple and irreversible way in order to show relationships among them (see chapter 3.1.2). Unfortunately, from the research, it emerges that European cultural content aggregators do not make visible the complex network of relationships among artworks. Often individual artworks lack the context of provenience and clear explanation about the presence of other related items. In most of the content aggregators, due to several restrictions in access tools (for both searching and browsing), the only way to access items sharing the same artistic period or location is by using related items in the surrogate page. This may generate an incomplete perception of the cultural collections heritage to general users and make it difficult the selection of an item.

The metadata usually available in a cultural content aggregator can reveal the following relationships among content: authors and artworks, cultural institutions and artworks or

places, artworks and other descriptive metadata (see fig. 5.2). Other possible relationships, such as those among different authors and works, require the production of other descriptive metadata by cultural institutions or end-users. These secondary descriptive metadata would be based on the interpretation of items by users and could be complementary to primary descriptive metadata.
European cultural content aggregators do not display networks revealing the relationships among artworks. In a few cases, they only display the distribution of items over a timeline.

Design strategies
From the research, it emerges that we can split the strategies to display connections among content into two levels: the general level, related to the overall set of surrogates, and the detail level, related to a single item. The aim of both the two levels is to foster the collection exploitability.
In the general level, the design of the user interface should make it visible the relationships among all the contents. Interactive tools should allow the creation of sets of surrogates according to their attributes, such as sets of items grouped by place, author, date and subject. At the general level, the

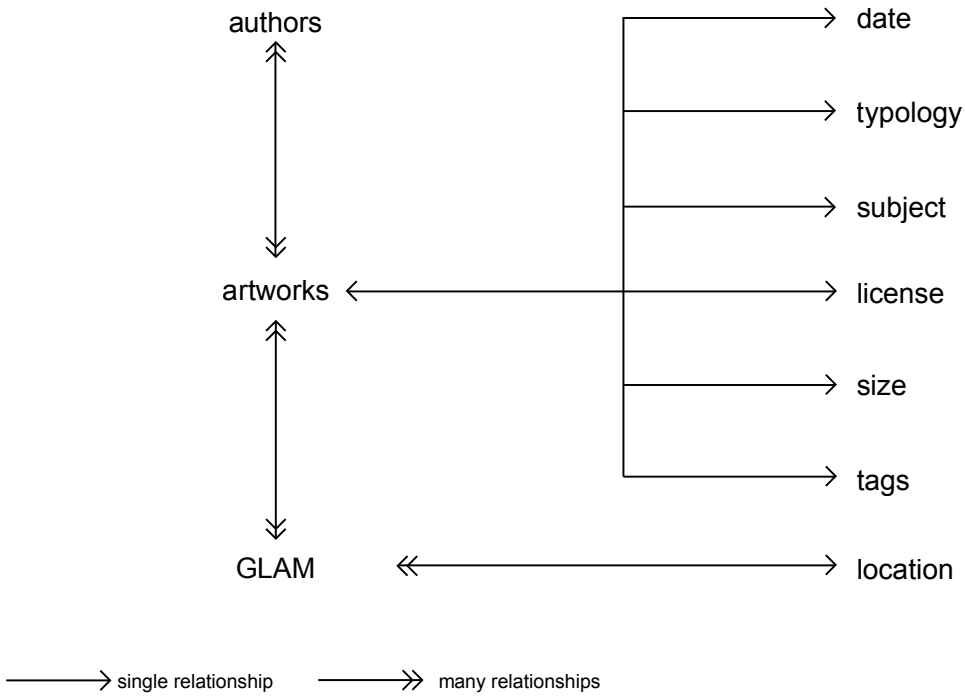


Fig. 5.2
Schema of the relationships among surrogate metadata.

interface might also provide users and GLAMs with navigation tools able to generate new connections among surrogates according to their interpretation.
In the detail level, the user interface should encourage pivoting. Interactive elements should allow users to move between sets of items sharing the same attributes. For instance, the metadata values can become a query for a new search. At the detail level, other possible design strategies for foster pivoting might be the use of surrogates over multiple pages, through the creation of user-generated content, and the integration of content coming from other related and relevant online sources.

User interface
In the creation of connections among content, the user interface should adopt some interactive elements that are already well-known in the field of data visualization and digital publishing (see tab. 5.2). With the data that are generally available, relevant networks can be generated between subjects that share artworks, authors and historical periods. The networks should feature as nodes only attribute with a smaller number of values, such as subjects, historical periods and mediums. Networks should also make visible the relevance of the nodes and the weight of the edges. The following example shows the characteristics of a subject-artwork network. In this network, the nodes are the subjects. The areas of the circles representing the nodes are equal to the number of artworks with a certain subject. The thickness of the lines (edge) connecting a couple of nodes is equal to the number of artworks sharing the same subject.
As access tool, multiple networks with the same specifics can be produced according to a fixed or customizable timeframe. it is appropriate to provide, especially for the homepage, an interactive introduction[6] showing how the interface works or textual instructions for use.
The detail page should feature several interactive elements fostering the collection exploration. In particular, to encourage pivoting the interface should highlight visual elements that bring users to other sets of surrogates with attributes in common with the selected one. The "related items" are some of the most common interactive elements to continue exploring the collection. Within them, it should be clearly defined as the attributes that they share with the selected surrogate. Visual or textual information should provide an overview of the related surrogates both in quantitative and

	General level	Detail level
Design strategies	Allow users to create sets of surrogates with similar characteristics	Allow users to visualise set of surrogates analogous to the selected one Allow user to generate content based on digital surrogates
Interface elements	Filtered lists/grids Networks	Related items Links to user-generated con-tent

Tab. 5.2
Design strategies and interface elements to create connections among content.

qualitative terms. Furthermore, the detail page should feature links to other internal pages featuring the surrogate and external pages, such as Wikipedia articles.

5.4 Support the use of surrogates

A user tends to use a digital surrogate if it is relevant and has high quality (see paragraph "Qualitative analysis" on chapter 4.3.1). We may split the actions undertaken with a digital surrogate in use and reuse. Use refers to the usage of the surrogate without user intervention, while reuse refers to the usage after a remaking process. The main purposes for using a surrogate are social sharing, artistic or scientific dissemination and culture promotion. The main purposes for reusing a surrogate are making art and producing physical surrogates (merchandising).
For supporting actions related to the use of digital surrogates it is necessary that cultural institutions upload surrogates and related metadata in accordance with user requirements. A relevant example of application supporting the access and use of surrogates is Google Images. The research has shown that many people use the tool provided by Google to download the images they need. This is not only due to the fact that it offers an enormous number of images coming from the entire web but also because of its interface features. Among these, there is the use of a dense grid of images and previews of related items. The user interface also provides users with filters (size, color, usage right, type and time) and an infinite scroll to load other search results. Recently, perhaps taking inspiration from other cultural content aggregators, Google Images has introduced also the collections, a page where users can store and manage their favourite images.

Unfortunately, from the research, it emerges that on Wikimedia Commons some cultural institutions contribute surrogates in low quality and with poor information. In some cases, the upload of low-quality images and internal documents might be interpreted as a use of the content aggregator as a mere tool for content storage.

Context

Releasing digital surrogates does not only mean adding free terms of use but also providing users with all the necessary information to understand surrogate characteristics and interface tools that foster the use and sharing. From the research, it emerges that surrogates must also have a high-quality in terms of technical characteristics of the digital file. If a cultural aggregator does not comply with this requirement, users will not be able to use it nor take it into consideration for future searches.

users essentially need information to understand the characteristics and the context of the provenance of the surrogate and a clear declaration of the terms of use (especially in cases where different countries have different licenses).

Some surrogate metadata, such as date, location and author, may be uncertain. However, it is necessary to make uncertainty explicit, although this could impact the use of a surrogate. Making uncertainty explicit can rise acceptance and trust in the cultural content aggregator (Windhager et al., 2019).

We can spit the audience of cultural aggregators according to the level of knowledge of the platform and the goal of the visit. Beginner users, who have few or no knowledge of the content aggregator, expect an easy-to-use tool to access collections, while expert users expect more advanced tools to access precise content. Users who have a clear idea of what to search expect to access content quickly, while users who have no idea need tools to explore content.

Design strategies

The cultural content aggregators should be considered as working tools. The design of the content aggregator should provide users with tools that allow and guide the use of digital surrogates at the collection and individual item level (see tab. 5.3).

At the level of the collection, the interface must provide users of bookmarking tools that allow the management of sets of surrogates. The aim of these tools is to allow users to or-

	Collection level	Surrogate level
Novice	Exploring content in a guided way (if he does not know what to look for) Accessing punctual results (if he knows what to look for)	Quick share and download
Expert	Exploring content in a flexible and independent way (if he does not know what to look for) Accessing punctual and de-tailed results (if he knows what to look for)	Save and organise surrogates in a personal page

Tab. 5.3
User requirements according to the knowledge about the cultural content aggregator

ganize content for future usage. Bookmarking tools should make it easy to add personal notes to single or multiple surrogates, add labels to multiple sets of items, aggregate elements and eventually provide options for download in batch.
At the level of the individual item, the interface must feature elements that allow users to quickly understand, share and download the surrogate. Related information may include where the artwork was used, such as temporary or permanent exhibitions and relevant author biographies. Other relevant use over relevant website might be included through the use of AI. Sharing tools should include buttons to share the item on the main social networks. Download options should include several surrogate sizes for use in both digital and paper support. At the level of the item, the interface should also provide other surrogates related by one of the three types of metadata or related by visual similarity (with the use of AI).

User interface
The user interface should provide users with tools that support them in managing digital surrogates. The managing tools should include interface features for bookmarking content and creating customised set of items. They could be arranged in groups, within taxonomies, or in filtered lists, within facet, in order to facilitate future access. Furthermore, the interface should provide users with easy-to-use download and sharing options.

5.5 Encourage users contribution

Cultural collections, as well as any multimedia content on the web, can reach a wide diffusion. Within cultural content aggregators, providing users with tools to create content based on digitised collections, such as articles, reports and collages, foster the spread of digital surrogates. A relevant example is Wikipedia and its sister projects: the community of Wikipedia contributors counts approximately 20 million users. The first version of Wikipedia was born in January 2001 as a spin-off of Nupedia, a free online encyclopaedia in English written and reviewed by experts. Jimmy Wales, CEO of the web portal who owned Nupedia, set the building of a publicly editable encyclopaedia as a goal. His aim was to give every single person free access to the sum of human knowledge. In 2003, Jimmy Wales founded the Wikimedia Foundation (WMF), a non-profit organisation with the goal to develop and monitor wiki-based projects and open content.
The experience of Wikipedia and other online communities shows that the (even small) contribution of several people to a relevant project allows the generation of a large amount of content. Users voluntarily group in a community and contribute to an online platform that, if it has valuable content, inspires trust and meets their needs.

Context
Users voluntarily group in a community and contribute to an online platform if it has valuable content, inspires trust and meets their needs (see paragraph “Communities” in chapter 3.2).
Nowadays few cultural content aggregators have a solid community of users. Indeed, many cultural institutions do not have a contact person who is in charge of managing relationships with the community. However, there are several initiatives, such as the GLAM-wiki initiative by the Wikimedia Foundation, and Europeana Network Association[7], that helps cultural institutions to share their resources through the engagement of communities.
Users need an information system that is rich in high-quality content and easy to use. Thus, users themselves may contribute, together with cultural institutions, to generate and validate content. If the user needs are met the individual end-user will be encouraged to contribute.
If the content aggregator does not comply with these requirements it remains a system in which only cultural insti-

tutions can generate content. This prevents the opportunity to have a collaborative platform that finds and fix errors and enhance content. The content aggregator should both allow the creation and editing of content by users but also prevent malicious actions.
There are many actions to be undertaken to encourage users to contribute beyond the graphical interface. Among these, there is the promotion of content via other communication channels the organisation of writing challenges.

Design strategies
Encouraging users to contribute and, consequently, build a community around the aggregator means providing them with tools able to generate content and increase the value of the platform. The content aggregator should allow users to perform both individual and collective actions (see fig. 5.3). Among the individual actions, there is the creation of content: the content aggregator should allow users to arrange personal sets of items to be kept private or public and also give them the opportunity to generate contents based on surrogates, such as articles and reviews, within the information system. Users should also be allowed to upload derivative works, which will obviously be subordinated to the original ones.

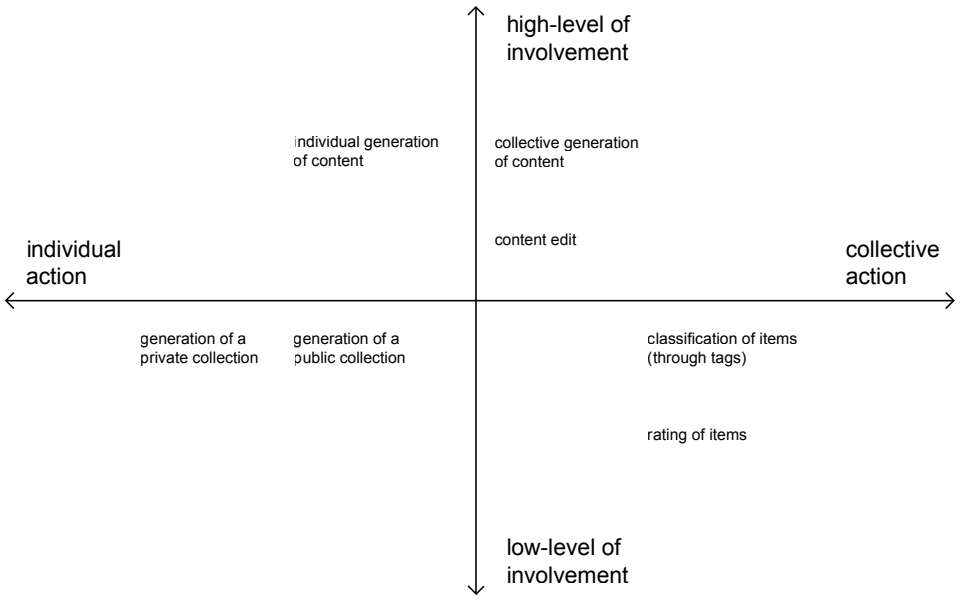


Fig. 5.3
Schema of the possible actions users can perform within a cultural content aggregator.

Among the collective actions, there are the creation and editing of content. In this case, the content aggregator may present the edit features as a way to suggest a content improvement and it should provide users with tools to contribute to the classification of digital surrogates, such as panels to add tags or categories. The aggregator should also allow users to rate surrogates and drop comments or feedback.

User interface

The implementation of interface tools to foster user contribution is challenging. It requires, in fact, the right balance between powerful interface tools and the ease of use of the cultural aggregator. The main interface features are the creation and editing of content, the update of the classification system and the release of comments and tags.

Tools for creating and editing content may include text editor and other interactive elements to allow users to contribute text, derivative works and other multimedia content. Tools for updating the classification may include forms to add tags and categories. Finally, tools for entering comments and feedback may include forms for commenting surrogates or give a feedback (such as like and dislike).

Notes

1. Link to OpenGLAM principles: https://openglam.org/principles/.
2. OpenGLAM initiative adopts the Open Definition, a document that states the precise meaning of "open" with respect to knowledge. The Open Definition essentially defines what knowledge is open, if anyone is free to access, use, modify, and share it, subject only to measures that preserve provenance and openness. The following is the link to the Open Definition: http://opendefinition.org/.
3. OpenGLAM CH website: https://glam.opendata.ch/.
4. Link to the results of the survey about OpenGLAM principles: https://openglam.org/2019/04/30/openglam-principles-ways-forward-to-open-access-for-cultural-heritage/.
5. Within facet systems that have a network schema, users may define relationships among classes of items.
6. The interactive introduction showing the main features of the web or mobile application is called "onboarding" in the user experience domain.
7. Europeana Network Association is a community of experts working in the field of digital heritage that aims to expand and improve access to Europe's digital cultural heritage.

The next Rembrandt,
by ING and Microsoft, in collaboration with advisors from TU Delft, The Mauritshuis and the Rembrandt House Museum, 2019. Website: https://www.nextrembrandt.com/

"The Next Rembrandt" is a 3D-printed painting made from the data of the artworks of Rembrandt (1606-1669). It was realised by using deep learning algorithms and facial recognition techniques. The researchers started the design process by gathering high-resolution scans of the complete collection of images of all 346 Rembrandt paintings. After classifying all the portrayed faces, they develop a software able to analyse the overall composition and painting materials, as well as specific features such as eyes, nose and mouth. The software learned how to create a Rembrandt face according to these features. Then, they moved from a 2D image to a 3D printed painting by using X-ray scans and training the system to distinct layers (canvas, ground layer and brushstrokes). Finally, they printed the portrait using thirteen layers of paint-based UV ink.
The technology developed is now used for the restoration of damaged masterpieces.

6

Conclusion

The final outcome of the thesis is a proposal of a series of guidelines for the design of cultural content aggregators, which I have defined through a stakeholder and user research.
The research showed that the audience of cultural content aggregators have a higher expertise than the one of digital archives. Thus, end-users of cultural aggregators need more powerful and direct interface tools. The research also showed that the OpenGLAM principles for the opening of digital collections are still little adopted today.
Cultural institutions particularly appreciate visual tools for monitoring the use of their collections online, while end-users choose visual tools only as a secondary access mode, after the search box.
From the research, it emerges that one of the fundamental aspects to increase the access and use of the digitised collections is to provide user interfaces with multiple access modes (search, browse and explore), as well as relevant connections among surrogates.
The design guidelines suggest designers to adopt some strategies to facilitate the access to surrogates and encourage the user-generated content. They can also provide designers with useful tips to redesign current cultural aggregators. The design guidelines are not conceived as point of arrival of the research on cultural content aggregators but they are a list of insights that can open to multiple fields of investigation related to the themes of archives and user engagement. This work reflects the current need of cultur-

al institutions to focus primarily on the needs of users and improve the accessibility of their digitised collections[1].

In a future scenario, cultural content aggregators may increase the dynamics of folksonomy and user-generated content. Cultural institutions will continue to digitise and re-digitise artworks thanks the introduction of more and more powerful digitisation machines and new digital formats. On cultural aggregators, physical artworks will be integrated with born-digital artworks (2D and 3D videos, digital images, creative coding). Thus, we might use the terms "digital surrogates" and "digital cultural objects". The living author of the artworks will interact with cultural aggregators and the digital artworks themselves might be released in open source[2].
Although I partially based my research on Wikipedia volunteers as a proxy with end-users, it was difficult to contact people using cultural content aggregators. This is also due to the weak relationships between the organizations managing the cultural aggregators and their end-users. However, as documented in the literature about the user research, even a small number of testers is enough to find out most of the usability issues and to obtain relevant insights. The second limit of the research is that the guidelines have not been validated through a redesign of GLAM Culture Hub or by asking GLAMs and the design community. However, each guideline is based on multiple evidences emerged from the literature review and from an in-depth empirical research. After a feedback from GLAMs and the design community, these design guidelines might change in form but not in substance.

Future works related to the thesis include the validation of the design guidelines through a new design project and the dissemination within the design community and GLAMs. Furthermore, I intend to expand the research on design practices in order to foster the spread and usage of digitised cultural collections. From a technical point of view, future works will include the optimisation and documentation of the scripts created to automate the data collection and processing.
In conclusion, the opening of cultural collections and the technological advancement is leading towards a reconceptualization of cultural content aggregators. The design of these platforms should be based on the real end-users needs. Finally, cultural content aggregators should not be considered as static searchable databases, but as dynamic research and dissemination tools.

Notes

1. Link to a Europeana article documenting future activities of the cultural content aggregator: https://ec.europa.eu/digital-single-market/en/news/ten-years-europeana-bringing-europes-cultural-heritage-digital-age.
2. Link to Reprogrammed art website - an art project about the release of interactive artworks in open source: http://www.reprogrammed-art.cc/.

References

A. Hearst, Marti. 2006. "Clustering versus Faceted Categories for Information Exploration." *Communications of the ACM - Supporting exploratory search* 49: 59–61.

Ahlberg, Christopher, and Ben Shneiderman. 2001. "Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays."

Anderson, Chris. 2008. *The Long Tail : Why the Future of Business Is Selling Less of More*. Rev. and u. New York: Hyperion.

Andrienko, Gennady, and Natalia Andrienko. 2007. "Coordinated Multiple Views: A Critical View." In *Proceedings - Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, CMV 2007, 72–74.

Bagnall, Rogers S. 2002. *Proceedings of the American Philosophical Society Alexandria: Library of Dreams*.

Bangor, Aaron, Philip T. Kortum, and James T. Miller. 2008. "An Empirical Evaluation of the System Usability Scale." *International Journal of Human-Computer Interaction* 24(6): 574–94.

Bates, Marcia. 2002. "Toward an Integrated Model of Information Seeking and Searching." *New Review of Information Behaviour Research* 3: 1–15.

———. 2007. "What Is Browsing - Really? A Model Drawing from Behavioural Science Research." *Information Research* 12.

Bearman, David, and Jennifer Trant. 2008. "Issues in Structuring Knowledge and Services for Universal Access to Online Science and Culture." In *Nobel Symposium "Virtual Museums and Public Understanding of Science and Culture,"* http://nobelprize.org/nobelfoundation/- publications/symposia/ns120-lectures/bearman.pdf.

Bekiari, Chryssoula, and Panos Constantopoulos. 2007. "Information Design for Cultural Documentation." In *Proceedings of the 9th DELOS Network of Excellence Thematic Workshop: Digital Repositories: Interoperability and Common Services, Crete, Greece: Heraklion*. http://delos-wp5.ukoln.ac.uk/dissemination/pdfs/constantopoulos-2.pdf.

Belkin, Nicholas J., Pier Giorgio Marchetti, and Colleen Cool. 1993. "BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval." *Information Processing & Management*: 325–44.

Belkin, Nicholas J., Robert N. Oddy, and Helen M. Brooks. 1982. "ASK for Information Retrieval: Part I. Background and Theory." *Journal of documentation* 38(2): 61–71.

Benhamou, Françoise, J. Manfred Holler, and Isidoro Mazza. 2013. *Handbook on the Economics of Cultural Heritage*. eds. Ilde Rizzo and Anna Mignosa.

Benjamin, Walter. 1936. "The Work of Art in the Age of Mechanical Reproduction." *Visual Culture: Experiences in Visual Culture*: 137–44.

Boot, Peter. 2014. "Distant Reading. Franco Moretti." *Digital Scholarship in the Humanities* 30(1): 152–54. https://doi.org/10.1093/llc/fqu010.

Booth, Ben. 1998. "Understanding the Information Needs of Visitors to Museums." *Museum Management and Curatorship* 17(2): 139–57.

Borowiecki, Karol Jan, and Trilce Navarrete. 2017. "Digitization of Heritage Collections as Indicator of Innovation." *Economics of Innovation and New Technology* 26(3): 227–46.

Botta, Massimo, Gillian Crampton Smith, and Giovanni Anceschi. 2006. *Design Dell'informazione: Tassonomie per La Progettazione Di Sistemi Grafici Auto-Nomatici*. Valentina Trentini.

Branchesi, Lida. 2016. Comunicare Il Museo Oggi : Dalle Scelte Museologiche Al Digitale : [Atti Del Convegno Internazionale Di Studi, Sapienza Università Di Roma, Dipartimento Di Storia Dell'Arte e Spettacolo, 18-19 Febbraio 2016]. Milano: Skira.

Brida, Juan Gabriel, Chiara Dalle Nogare, and Raffaele Scuderi. 2016. "Frequency of Museum Attendance: Motivation Matters." *Journal of Cultural Economics* 40(3): 261–83.

Buchanan, Brian. 1979. Theory of Library Classification. London: Bingley.

Burns, Jasmine E. 2017. "The Aura of Materiality: Digital Surrogacy and the Preservation of Photographic Archives." *Art Documentation: Journal of the Art Libraries Society of North America* 36(1): 1–8.

Cangiano, Serena et al. 2017. "Community Digest: Using Data to Visualize Wikipedia Knowledge Gaps; News in Brief." Wikimedia Foundation Blog. http://repository.supsi.ch/9196/.

Caplan, Priscilla. 2003. *Metadata Fundamentals for All Librarians*. Chicago: American Library Association.

Card, Stuart K. 1999. *Readings in Information Visualization: Using Vision to Think*. San Francisco, Calif: Morgan Kaufmann Publishers.

Cassella, Maria. 2013a. "La Gestione Dei Diritti

Nei Progetti Di Digitalizzazione: Il Pubblico Dominio e Le Opere Orfane." JLIS.it 4(2): 223–54.
———. 2013b. "Rights Management in Digitization Projects: Public Domain and Orphan Works." JLIS.it 4(2): 223–54.
Casson, Lionel. 2002. Libraries in the Ancient World. Yale University Press.
Chen, Ko-le, Marian Dörk, and Martyn Dade-Robertson. 2014. "Exploring the Promises and Potentials of Visual Archive Interfaces."
Conway, Paul. 2014. 15 *Archival Science Digital Transformations and the Archival Nature of Surrogates*.
Cook, Terry. 2001. "Archival Science and Postmodernism: New Formulations for Old Concepts." *Archival Science* 1(1): 3–24. https://doi.org/10.1007/BF02435636.
Coughlin, Daniel, Mark C. Campbell, and Jim Jansen. 2013. 50 *Proceedings of the American Society for Information Science and Technology Measuring the Value of Library Content Collections*.
Currall, James, Michael Moss, and Susan Stuart. 2005. "What Is a Collection." *Archivaria* 58: 131–146.
Davis, Douglas. 1995. "The Work of Art in the Age of Digital Reproduction (An Evolving Thesis: 1991-1995)." *Leonardo* 28(5): 381.
Deal, Laura. 2015. "Visualizing Digital Collections." T*echnical Services Quarterly* 32(1): 14–34.
Dörk, Marian, Sheelagh Carpendale, and Carey Williamson. 2011. "The Information Flaneur: A Fresh Look at Information Seeking." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1215–24.
Dörk, Marian, Christopher Pietsch, and Gabriel Credico. 2017a. "One View Is Not Enough: High-Level Visualizations of a Large Cultural Collection." *Information Design Journal* 23: 39–47.
———. 2017b. "One View Is Not Enough." Information Design Journal 23(1): 39–47.
Dörk, Marian, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. 2012. "Pivotpaths: Strolling through Faceted Information Spaces." *IEEE transactions on visualization and computer graphics* 18(12): 2709–18.
Drucker, Johanna. 2013. "Performative Materiality and Theoretical Approaches to Interface." DHQ: Digital Humanities Quarterly 7(1).
Fantoni, Silvia Filippini, Rob Stein, and Gray Bowman. 2012. "Exploring the Relationship between Visitor Motivation and Engagement in Online Museum Audiences." In *Museums and the Web*.
Farber, Allison, and Paul Radensky. 2008. "Living Museum®: Supporting the Creation of Quality User-Generated Content." In Museums and the Web, eds. J. Trant and D. Bearman. Toronto: Archives & Museum Informatics. http://www.archimuse.com/mw2008/papers/farber/farber.html.
Farrauto, Luigi, and Giovanni Profeta. 2014. "Tracce Digitali." Progetto grafico (26). http://repository.supsi.ch/10480/.
Fidel, Raya. 2012a. Human Information Interaction : An Ecological Approach to Information Behavior. Cambridge, Mass [etc.: MIT Press.
———. 2012b. Human Information Interaction: An Ecological Approach to Information Behavior Human Information Interaction: An Ecological Approach to Information Behavior. MIT Press.
Fiorio, Maria Teresa. 2011. Il Museo Nella Storia: Dallo Studiolo Alla Raccolta Pubblica. Milano: Bruno Mondadori.
Ford, Heather et al. 2018. "Beyond Notification: Filling Gaps in Peer Production Projects." New Media and Society 20(10): 3799–3817.
Frayling, Christopher. 1994. Research in Art and Design (Royal College of Art Research Papers, Vol 1, No 1, 1993/4). London: Royal College of Art. http://researchonline.rca.ac.uk/384/.
Fu, Bo, Natalya F. Noy, and Margaret-Anne Storey. 2013. "Indented Tree or Graph? A Usability Study of Ontology Visualization Techniques in the Context of Class Mapping Evaluation." In The Semantic Web - ISWC 2013, eds. Harith Alani et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 117–34.
Ginsburgh, Victor, and Jan C. van Ours. 2003. "Expert Opinion and Compensation: Evidence from a Musical Competition." American Economic Review 93: 289–96.
Glinka, Katrin, Christopher Pietsch, and Marian Dörk. 2017. "Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections." Digital Humanities Quarterly 11.
Gnoli, Claudio, Vittorio Marino, and Luca Rosati. 2006. Organizzare La Conoscenza. Dalle Biblioteche All'architettura Dell'informazione per Il Web. Tecniche nuove.
Gordon Lanning, Robbyn, and Jonathan B. Bengtson. 2016. "Traces of Humanity: Echoes

of Social and Cultural Experience in Physical Objects and Digital Surrogates in the University of Victoria Libraries" ed. Aaron Mauro. Cogent Arts & Humanities 3(1): 1163042. https://doi.org/10.1080/23311983.2016.1163042.

Guerrini, Mauro, and Rosa Maiello. 2010. "Si Fa Presto a Dire 'Biblioteche Digitali'. Un Confronto Tra Google Book Search Ed Europeana." La Bibliofilía 112(1): 77–94. http://www.jstor.org/stable/26198871.

Gwen, Glazer. 2010. "Digitizing Hidden Collections in Public Libraries." American Library Association. http://www.ala.org/aboutala/sites/ala.org.aboutala/files/content/oitp/publications/oitpperspectives/oitp_perspectives_ju.pdf.

Herring, Susan C. 2011. "Where Are the Women in Wikipedia? - Room for Debate." The New York Times. https://www.nytimes.com/roomfordebate/2011/02/02/where-are-the-women-in-wikipedia (September 17, 2019).

Herseni, Johannes et al. 2018. "Die Bibliografischen Daten Der Deutschen Nationalbibliothek Entfalten." Dialog mit Bibliotheken 1: 11–16.

Hill, Timothy, Valentine Charles, Juliane Stiller, and Antoine Isaac. 2016. "'Searching for Inspiration': User Needs and Search Architecture in Europeana Collections." Proceedings of the Association for Information Science and Technology 53(1): 1–7. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pra2.2016.14505301043.

Hinrichs, Uta, Stefania Forlini, and Bridget Moynihan. 2015. "Speculative Practices: Utilizing InfoVis to Explore Untapped Literary Collections." IEEE Transactions on Visualization and Computer Graphics 22: 1.

Hodge, Gail M. 2000. 6 D-lib Magazine - DLIB Best Practices for Digital Archiving.

Howes, Deborah Seid. 2007. "Why the Internet Matters: A Museum Educator's Perspective." In The Digital Museum. A Think Guide., ed. American Association of Museums. Washington, DC: John Strand, 67–78.

Hughes, Lorna M. 2011. "Introduction: The Value, Use and Impact of Digital Collections." In Evaluating and Measuring the Value, Use and Impact of Digital Collections, ed. LornaEditor Hughes. Facet, 1–10.

IDEO.org. 2015. The Field Guide to Human-Centered Design.

Jacobsen, Trond, Ricardo Punzalan, and Margaret Hedstrom. 2013. "Invoking 'Collective Memory': Mapping the Emergence of a Concept in Archival Science." Archival Science 13.

Johnson, Peter, and Barry Thomas. 1998. "The Economics of Museums: A Research Perspective." Journal of cultural economics 22(2–3): 75–85.

Kapsalis, Effie. 2016. The Impact of Open Access on Galleries, Libraries, Museums, & Archives.

Kräutli, Florian. 2016. "Visualising Cultural Data. Exploring Digital Collections Through Timeline Visualisations." Royal College of Art.

Kreiseler, Sarah, Viktoria Brüggemann, and Marian Dörk. 2017. "Tracing Exploratory Modes in Digital Collections of Museum Web Sites Using Reverse Information Architecture." First Monday 22(4). https://firstmonday.org/ojs/index.php/fm/article/view/6984/6090%0A.

Krug, Steve. 2000. Don't Make Me Think. San Francisco, California: New Riders Press.

———. "Don't Make Me Think!: A Common Sense Approach to Web Usability." https://www.researchgate.net/publication/234779599_Don%27t_Make_Me_Think_A_Common_Sense_Approach_to_Web_Usability (September 17, 2019).

Kumar, Krishna. 1991. Theory of Classification. Vikas Publishing House. https://books.google.ch/books?id=lMeHcgAACAAJ.

Lee, Hur-Li. 2005. 75 Library Quarterly The Concept of Collection from the User's Perspective.

Lehane, Richard. 2012. "Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-Prospect Browsing." Archives and Manuscripts 40: 110–12.

Lin, Yi-Ling, Peter Brusilovsky, and Daqing He. 2016. "Finding Cultural Heritage Images through a Dual-Perspective Navigation Framework." Information Processing & Management 52(5): 820–39.

Lopatovska, Irene, Iris Bierlein, Heather Lember, and Eleanor Meyer. 2013. "Exploring Requirements for Online Art Collections." In Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries, 111.

Manovich, Lev. 1999. "Database as Symbolic Form." Convergence 5(2): 80–99. https://doi.org/10.1177/135485659900500206.

———. 2002. "Data Visualisation as New Abstraction and Anti-Sublime." Small tech: The culture of digital tools 22: 1.

———. 2007a. "Cultural Analytics: Analysis and

Visualization of Large Cultural Data Sets." Manovich.net 30.
———. 2007b. "What Comes after Remix." Remix Theory 10.
———. 2012a. "How to Compare One Million Images?" In Understanding Digital Humanities, London: Palgrave Macmillan UK, 249–78. http://link.springer.com/10.1057/9780230371934_14 (September 17, 2019).
———. 2012b. "Media Visualization: Visual Techniques for Exploring Large Media Collections." The international encyclopedia of media studies.
Marchionini, Gary. 2006. "Exploratory Search: From Finding to Understanding." Communications of the ACM 49(4): 41–46.
Marini Clarelli, Maria Vittoria. 2007. Che Cos'è Un Museo. 4a rist. Roma: Carocci.
Masnizah, Mohd. 2012. "Evaluation of an Interactive Topic Detection and Tracking Interface."
Matusia, Krystyna K., and Tamara K. Johnston. 2014. "Digitization for Preservation and Access: Restoring the Usefulness of the Nitrate Negative Collections at the American Geographical Society Library." The American Archivist 77(1): 241–69.
McCarty, Willard, and Matthew Kirschenbaum. 2003. "Institutional Models for Humanities Computing." Literary and Linguistic Computing 18(4): 465–89. https://doi.org/10.1093/llc/18.4.465.
Meho, Lokman, and Helen R. Tibbo. 2003. 54 Journal of the American Society for Information Science and Technology Modeling The Information- Seeking Behavior of Social Scientists: Ellis's Study Revisited.
Meirelles, Isabel. 2011. "Visualizing Data: New Pedagogical Challenges." In Selected Readings of the 4th Information Design International Conference, 73–83. http://openresearch.ocadu.ca/id/eprint/1784/.
Menichinelli, Massimo. 2007. "OpenP2PDesign." Politecnico di Milano. http://www.openp2pdesign.org/.
Menichinelli, Massimo, Giovanni Profeta, and Serena Cangiano. 2020. 2CO COmmunicating COmplexity Visualising Open Communities. Guidelines from Three Case Studies. http://repository.supsi.ch/10966/.
Meyer, Eric T. 2009. Final Report to JISC on the Usage and Impact Study of JISC-Funded Phase 1 Digitisation Projects and the Toolkit for the Impact of Digitised Scholarly Resources TIDSR. Oxford.
Monaci, Sara. 2005. Il Futuro Nel Museo : Come i Nuovi Media Cambiano l'esperienza Del Pubblico. Milano: Guerini Studio.
Moretti, Franco. 2013. Distant Reading. Verso.
Morgan, David L. 2014. "Pragmatism as a Paradigm for Social Research." Qualitative Inquiry 20(8): 1045–53. https://doi.org/10.1177/1077800413513733.
Morville, Peter. 2005. Ambient Findability: What We Find Changes Who We Become.
Navarrete, Trilce, and Karol Jan Borowiecki. 2015. ACEI Working Paper Series Change In Access After Digitization: Ethnographic Collections In Wikipedia Change In Access After Digitization: Ethnographic Collections In Wikipedia. https://en.wikipedia.org/wiki/Wikipedia. (October 2, 2019).
———. 2016. "Changes in Cultural Consumption: Ethnographic Collections in Wikipedia." Cultural Trends 25(4): 233–48.
Nicholas, David. 2009. Assessing Information Needs in the Age of the Digital Consumer. Third ed. ed. Eti Herman. London [etc.: London [etc.] : Routledge, 2009.
Organisciak, Piotr. 2010. "Why Bother? Examining the Motivations of Users in Large-Scale Crowd-Powered Online Initiatives." University of Alberta. https://era.library.ualberta.ca/items/65ae538e-45ec-4bc2-a915-06db38844412/view/ca5e10e9-376d-4aae-a660-3af938de4e6d/ThesisOrganisciak-08-2010v2.pdf.
Panetto, Hervé, and Joe Cecil. 2013. "Information Systems for Enterprise Integration, Interoperability and Networking: Theory and Applications." Enterprise Information Systems 7(1): 1–6. https://doi.org/10.1080/17517575.2012.684802.
Paqua, Megan. 2018. "Beyond Digitization: Planning for Open Access Collections." Pratt. Master of Science in Museums and Digital Culture. https://museumsdigitalculture.prattsi.org/beyond-digitization-planning-for-open-access-collections-fb-7325163f5d.
Pearce-Moses, Richard. 2005. A Glossary of Archival and Records Terminology.
Phillips, Heather. 2010. 2010 Library Philosophy and Practice The Great Library of Alexandria?
Pomerantz, Jeffrey. 2015. Metadata. Cambridge MA: The MIT Press.
Pomerantz, Jeffrey, and Robin Peek. 2016. "Fifty

Shades of Open." First Monday 21.
Powell, Sarah. 2016. "Enriching Our Online Collections - Managing Image Rights for Digital Surrogate." Auckland Museum blog. https://www.aucklandmuseum.com/discover/stories/blog/2016/enriching-our-online-collections-rights-management (May 6, 2019).
Profeta, Giovanni. 2018. "Visualizzare Collezioni Culturali Aperte. Metodi per l'esplorazione Dei Contenuti Culturali Su Wikipedia." In FRID 2017: Fare Ricerca in Design. Sul Metodo/Sui Metodi: Esplorazioni per Le Identità Del Design, ed. Raimonda Riccini. Mimesis, 453–63. http://repository.supsi.ch/10964/.
Pruulmann-Vengerfeldt, Pille, and Agnes Aljas. 2011. "Digital Cultural Heritage – Challenging Museums, Archives and Users." Journal of Ethnology and Folkloristics 3.
Ramirez, Artemio, Joseph Walther, Judee Burgoon, and Michael Sunnafrank. 2002. "Information-Seeking Strategies, Uncertainty, and Computer-Mediated Communication Toward a Conceptual Model." Human Communication Research - HUM COMMUN RES 28: 213–28.
Raven, Peter H., Brent Berlin, and Dennis E. Breedlove. 1971. "The Origins of Taxonomy." Science 174(4015): 1210–13. http://www.jstor.org/stable/1732886.
Reinoso, Antonio J., Jesus M. Gonzalez-Barahona, Felipe Ortega, and Gregorio Robles. 2008. "Quantitative Analysis and Characterization of Wikipedia Requests." In WikiSym 2008 - The 4th International Symposium on Wikis, Proceedings,.
Rice, Ronald E., Maureen McCreadie, and Shan-Ju Chang. 2002. 7 Accessing and Browsing Information and Communication.
Rick, Szostak. 2016. "Synthetic Classification of Museum Artifacts Using Basic Concepts." Museums and the Web. https://mw2016.museumsandtheweb.com/paper/synthetic-classification-of-museum-artifacts-using-basic-concepts/.
Rimmer, Jon et al. 2008. "An Examination of the Physical and the Digital Qualities of Humanities Research." Information Processing & Management 44: 1374–92.
Roberto, Frankie. 2008. "Exploring Museum Collections On-Line: The Quantitative Method." In Museums and the Web, eds. J. Trant and D. Bearman. Toronto: Archives & Museum Informatics.
Roberts, Lisa. 1997. From Knowledge to Narrative: Educators and the Changing Museum. Washington DC: Smithsonian Institution Press.
Satija, M. P., and Daniel Martinez-Avila. 2015. 3 Journal of Information Science Theory and Practice Features, Functions and Components of a Library Classification System in the LIS Tradition for the e-Environment.
Shein, Cyndi, and Emily Lapworth. 2016. "Say Yes to Digital Surrogates: Strengthening the Archival Record in the Postcustodial Era." Journal of Western Archives 7(1).
Shneiderman, Ben. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In Proceedings 1996 IEEE Symposium on Visual Languages, 336–43.
Sotirova, Kalina et al. 2012. "Digitization of Cultural Heritage–Standards, Institutions, Initiatives." https://www.academia.edu/2855210/Digitization_of_Cultural_Heritage_Standards_Institutions_Initiatives (October 13, 2019).
Spence, Robert. 2007. Information Visualization. 2nd, fully ed. Harlow, UK [etc.: Pearson Education/Prentice Hall.
Stiller, Juliane. 2012. "A Framework for Classifying Interactions in Cultural Heritage Information Systems." International Journal of Heritage in the Digital Era 1(1_suppl): 141–46. https://doi.org/10.1260/2047-4970.1.0.141.
———. 2014. "From Curation to Collaboration. A Framework for Interactions in Cultural Heritage Information Systems." Humboldt-Universität.
Terras, Melissa. 2015. "Opening Access to Collections: The Making and Using of Open Digitised Cultural Content." Online Information Review (Special Issue on "Open Access: Redrawing the Landscape of Scholarly Communication"). https://discovery.ucl.ac.uk/id/eprint/1469561/1/MelissaTerras_OpeningAccess_OIR.pdf.
Toms, Elaine. 2000. "Understanding and Facilitating the Browsing of Electronic Text." International Journal of Human-Computer Studies 52: 423–52.
Tufte, Edward Rolf. 2001. The Visual Display of Quantitative Information. 2nd ed. Cheshire, Connecticut: Graphics Press.
Ushizima, Daniela, Lev Manovich, Todd Margolis, and Jeremy Douglass. 2012. "Cultural Analytics of Large Datasets from Flickr." : 30–34.
Wallace, Andrea, and Deazley Ronan. 2016. Introduction to Display At Your Own Risk: An

Experimental Exhibition of Digital Cultural Heritage. Cambridge MA: Cambridge University Press. https://displayatyourownrisk.org/.
Whitelaw, Mitchell. 2012. “Towards Generous Interfaces for Archival Collections.” International Council on Archives Congress 2012.
———. 2015. “Generous Interfaces for Digital Cultural Collections.” Digital Humanities Quarterly 9.
———. 2016. “Representing Digital Collections.” In Performing Digital, Routledge, 97–116.
Wiener, Norbert. 2014. Cybernetics or Control and Communication in the Animal and the Machine. 2nd pbk. e. Cambridge, Mass: Cambridge, Mass. : M.I.T. Press, [2014].
Williams, David. 2010. “A Brief History of Museum Computerization.” In Museum in a Digital Age, ed. Ross Parry. London: Routledge.
Wilson, Tom. 1999. “Models in Information Behaviour Research.” Journal of Documentation 55.
Windhager, Florian et al. 2019. “Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges.” IEEE Transactions on Visualization and Computer Graphics 25(6): 2311–30.
Windhager, Florian, Saminu Salisu, and Eva Mayr. 2019. “Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections.” Informatics 6(3): 29.
Xie, Hong. 2000. “Shifts of Interactive Intentions and Information-Seeking Strategies in Interactive Information Retrieval.” JASIS 51: 841–57.
Zeng, Marcia Lei, and Qin Jian. 2008. Metadata. New York: Neal-Schuman Publishers.
Zoller, Gabriela, and Katie DeMarsh. 2013. “For the Record: Museum Cataloging from a Library and Information Science Perspective.” Art Documentation: Journal of the Art Libraries Society of North America 32(1): 54–70. https://doi.org/10.1086/669989.

Glossary

Aggregator

An aggregator is an information system that collects and manages metadata from multiple data providers. It may also include multimedia content, such as images, videos and audio files. Within an aggregator, contents are stored and retrieved according to their metadata.

Artificial intelligence

Artificial intelligence (AI) is the capability of smart machines and software to perform tasks that typically require human intelligence. Artificial intelligence involves several technologies and processes, including machine learning and deep learning. Among the applications of AI, there are speech recognition, machine vision and natural language generation.

Catalogue

A catalogue is a systematic organisation of multiple items with some characteristics in common. Within the cultural heritage sector, this term refers to the list of cultural objects owned by a cultural institution.

Category

Category refers to a term that frames an item belonging to a catalogue within a specific thematic area. Categories may have a different degree of specificity according to the complexity of the catalogue of items, which can be displayed as a tree.

Classification

Classification refers to the procedure of grouping items according to specific characteristics. It aims to allow users to find items, identify duplicates and understand the context of an item quickly. The classification requires the creation of metadata and the use of a specific type of notation.

GLAM

GLAM is an acronym that stands for Galleries, Libraries, Archives and Museums. This acronym is generally used in the field of digital humanities to refer to cultural institutions that own digitised cultural collections. It is also adopted by open community and non-profit organisations to refer to cultural institutions which release their digitised collections under open licenses.

Indexing

Indexing refers to the generation of a catalogue of indexes starting from an extensive set of content. The index generation is an activity generally performed through computational techniques. It allows the rapid retrieval of content based on the submission of a query, within a computer-based information system.

Interoperability

Interoperability is the ability of physically disconnected information systems to perform specific tasks together. In the context of digitised collections, interoperability refers to the correct interpretation of data and functionality among different computer-based applications.

Database

A database is a structured set of data generally handled by a computer. It allows users to submit a query in order to get a list of pertinent records. In computer science, databases can store several types of data through different database models (such as relational and non-relational data model). In the context of digitised collections, databases store metadata of the digitised objects.

Digitisation

Digitisation is the process of converting item from a physical format into a digital one. In the cultural content field, the main technologies to digitise artworks are photography, and 2D and 3D scanners. The digitisation process also involves the creation of metadata. Both software and humans can perform this task.

Facet

A facet is a semantic category, either general or subject-specific, that is combined with others to create the full classification entry. The generation of a facet is based on a specific characteristic of the items that constitute a catalogue.

Folksonomy

Folksonomy refers to a user-generated system of classifying and organising content into different categories by the use of metadata. Folksonomy usually takes place within web-based information systems, and it grounds on the generation of tags or categories by end-users.

Linked Open Data

Linked Open Data (LOD) is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found and links between related resources can be made.

Metadata

Metadata stands for something that is beyond the data related to an item, at a higher level of abstraction. There are four types of metadata: descriptive, structural, administrative and usage data (or paradata). Descriptive metadata concerns the creation of an item. Structural metadata refer to its structure and composition. Administrative metadata are about the provenance and the maintenance of an item, while usage metadata refer to its usage.

Open license

An open license is a document that specifies the permission, granted by the author of artwork, to access, use and redistribute his work with few or no restrictions. In the cultural sector, the most known open licenses are Creative Commons licenses.

Surrogate

Surrogate refers to an item that can be used as a representative of original items. It can be either a physical or digital object. Usually, cultural institutions use surrogates of the cultural object to preserve or to foster access to the original ones.

Tag

Tag refers to a term that describes an item belonging to a catalogue. An item can have multiple tags with a wide range of degrees of specificity. Since tags are not structured, they usually serve as a secondary way to classify a catalogue of items, after the categories.

Visual model

A visual model is a combination of visual elements and layout that aim to represent data. In other words, it is an archetype of a chart (such as bar chart and Sankey diagram). In the cultural sector, the most used visual models include timelines, maps, networks and plots.

Wiki

Wiki is a collaborative website where users can edit and create content even without having grant permission as an administrator. Most of the existing Wikies use a simplified mark-up language and do not have a rigid administration hierarchy, either a defined structured of pages. This allows the constant evolution of the website according to the needs of the users.

Appendix

Case studies analysis

I analysed 14 cultural content aggregators that allow the user to interact with digital surrogates, in order to identify the most common classification systems and interface solutions used by the cultural information systems.

Quantitative analysis

index	name	website
1	Archives Portal Europe	https://www.archivesportaleurope.net/
2	Culture Grid	http://www.culturegrid.org.uk/
3	Deutsche Digitale Bibliothek	https://www.deutsche-digitale-bibliothek.de/
4	EFG - The European Film Gateway	https://www.europeanfilmgateway.eu/
5	EUscreen	http://www.euscreen.eu/
6	Europeana	https://www.europeana.eu/portal/en
7	Hispana	http://hispana.mcu.es/es/inicio/inicio.do
8	Kultur Pool	http://www.kulturpool.at
9	Moteur Collections	http://www.culture.fr/Ressources/Moteur-Collections
10	SearchCulture	https://www.searchculture.gr/aggregator/portal/?language=en
11	Swiss National Library	https://www.helveticarchives.ch
12	The European Library	http://www.theeuropeanlibrary.org
13	The National Library of Finland	https://www.kansalliskirjasto.fi/en
14	Wikimedia Commons	https://commons.wikimedia.org/wiki/Main_Page

Tab. 1
European cultural content aggregators analysed.

index	title	date	author	tags	categories	subject	typology	colour/s	description	GLAM
1	x	x	-	x	-	-	-	-	-	x
2	x	-	-	-	-	x	-	-	x	x
3	x	x	-	-	-	-	-	-	x	x
4	x	x	x	x	x	-	x	x	-	x
5	x	x	-	x	-	-	x	x	x	x
6	x	x	x	-	x	-	x	x	x	x
7	x	x	x	-	-	x	-	-	-	x
8	x	x	-	-	-	x	x	-	-	x
9	x	x	x	-	-	x	x	-	x	x
10	x	x	-	-	x	-	x	-	-	x
11	x	x	x	-	x	-	x	-	x	x
12	x	x	x	-	-	x	x	-	x	x
13	x	x	x	-	-	x	x	-	-	x
14	x	x	x	-	x	-	x	-	x	x

Tab. 2
Metadata available on cultural content aggregators analysed.

index	total surrogates	GLAMs contributing	Google page rank	Pages indexed on Google	daily visits (homepage)
1	274.306.471	14	7	122.000	520.000
2	778.000	31	6	4.040	5.600.000
3	9.491.371	13	8	505.000	196.000
4	711.000	38	7	20.400	1.340.000
5	60.000	34	7	6.350	815.000
6	59.000.000	236	9	1.800.000	70.000
7	8.400.000	226	0	15.400	60.000
8	227.000	11	0	329.000	1.253.161
9	7.500.000	9	0	664.000	100.000
10	488.000	59	0	16.200	1.000
11	10.000	2	0	15.700	1.000
12	175.000.000	48	10	45.200	695.000
13	820.000	6	7	10.300	326.000
14	>1.000.000	>300	7	27.200.000	300.000

Tab. 3
Statistics of the cultural content aggregators analysed.

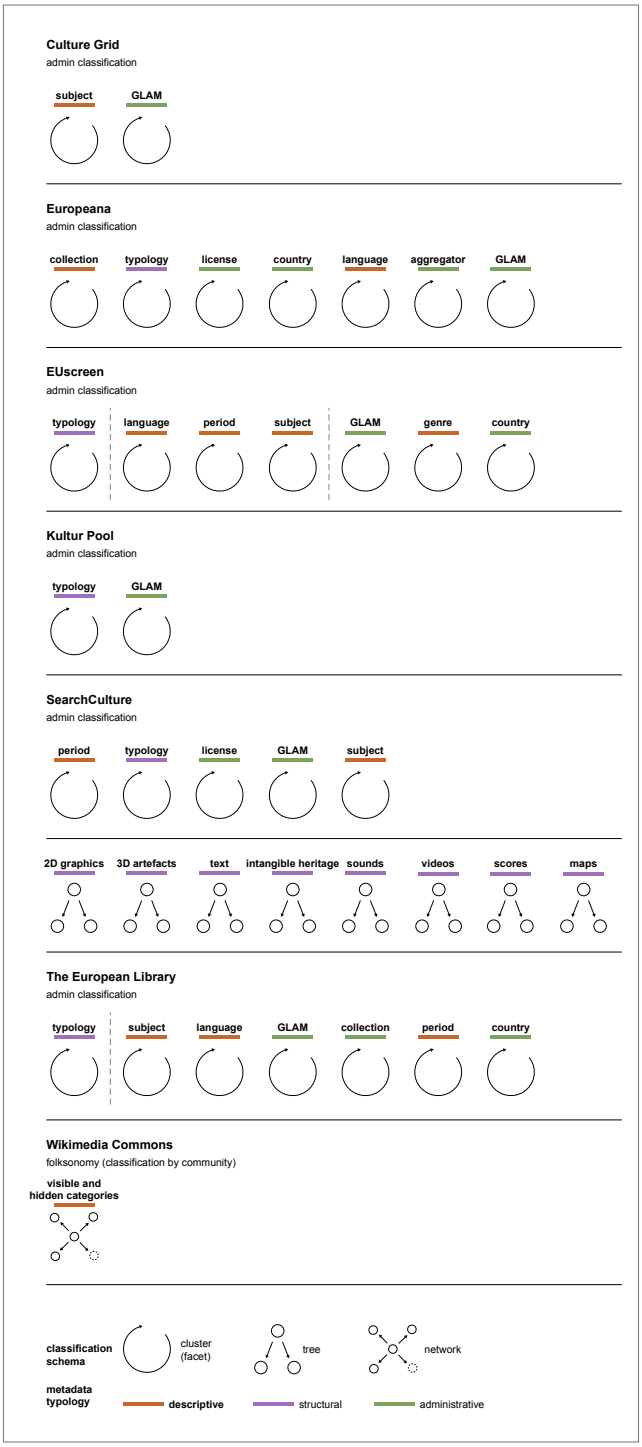


Fig. 1
Classification system analysis.

Information architecture analysis

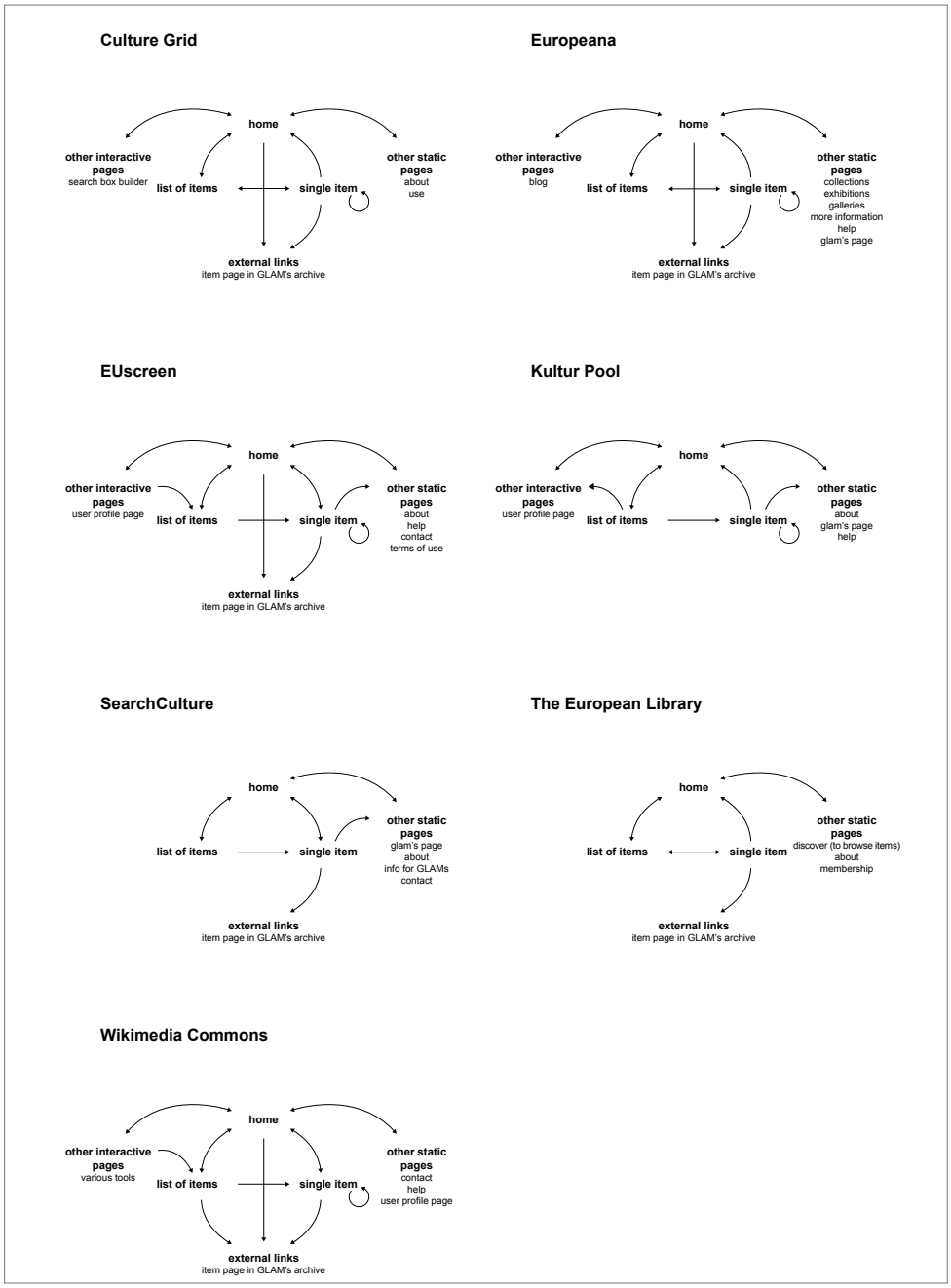


Fig. 2
Information architecture analysis.

Access modes analysis

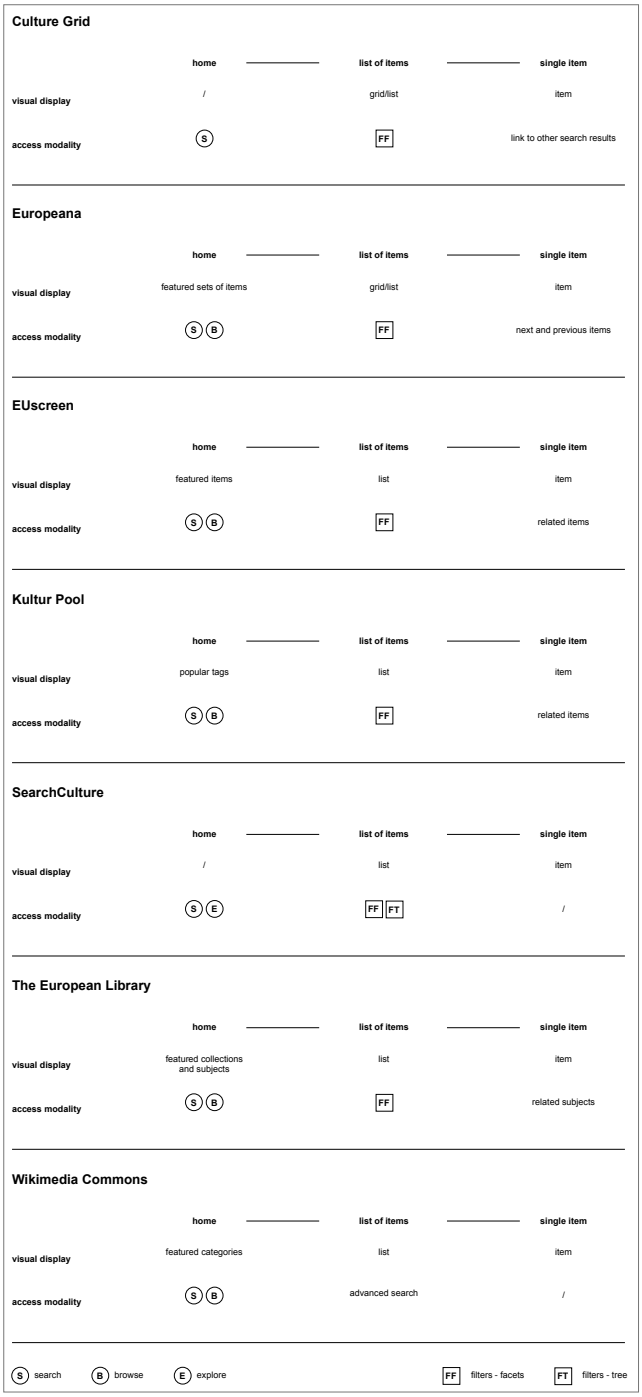


Fig. 3
Access modes analysis.

Layout analysis
Aggregator: Culture Grid.
By: Collections Trust.
Website: http://www.culturegrid.org.uk/

Fig. 4
Culture Grid homepage.

Fig. 5
Culture Grid layout analysis.

Aggregator: Europeana.
Funded by: Europeana Foundation.
Website: https://www.europeana.eu/

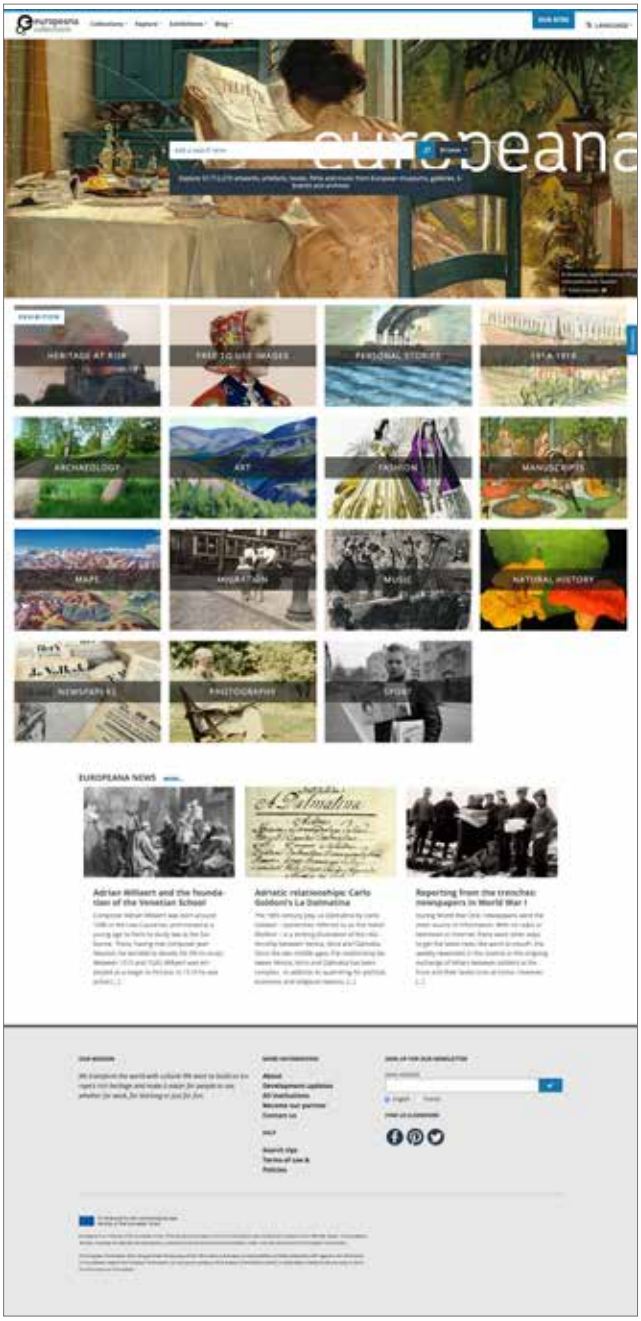


Fig. 6
Europeana homepage.

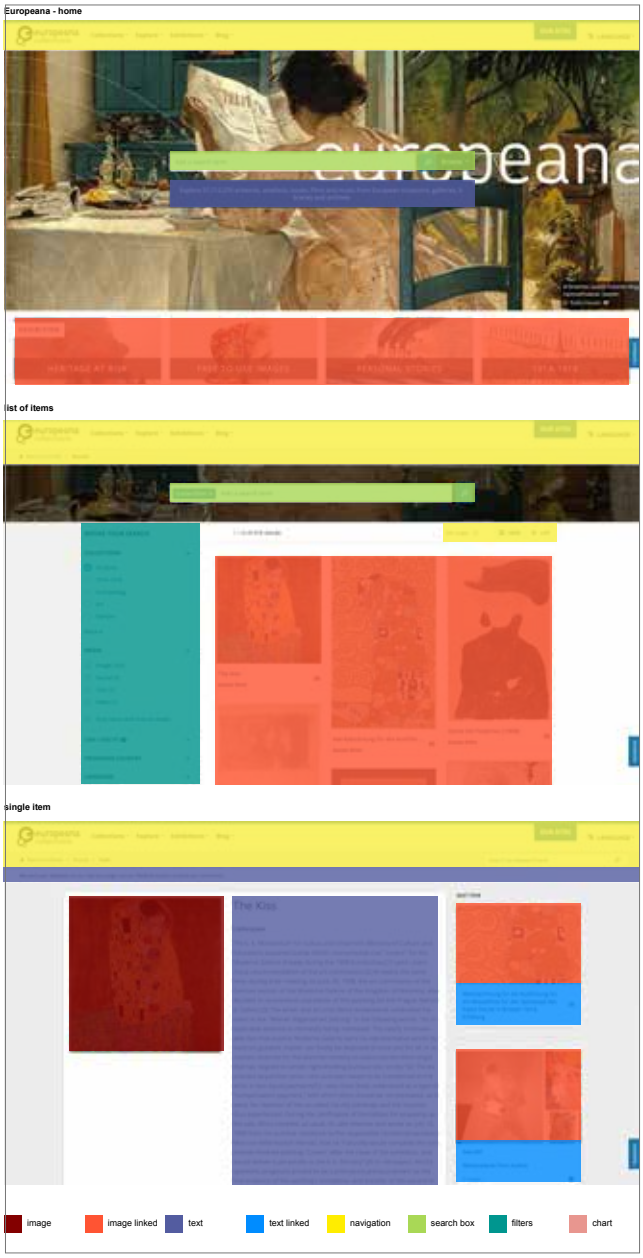


Fig. 7
Europeana layout analysis.

Aggregator: EUscreen.
Funded by: European Commission within the eContentplus programme.
Website: http://www.euscreen.eu/

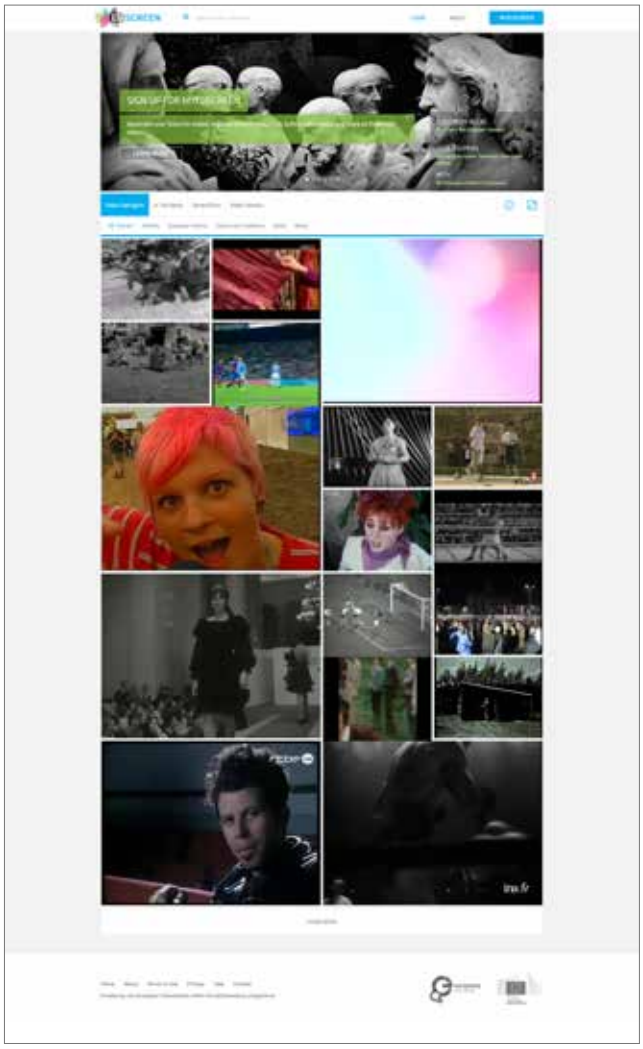


Fig. 8
EUscreen homepage

Fig.9
EUscreen layout analysis.

Aggregator: Kultur Pool.
Funded by: German Chancellery.
Website: http://www.kulturpool.at/

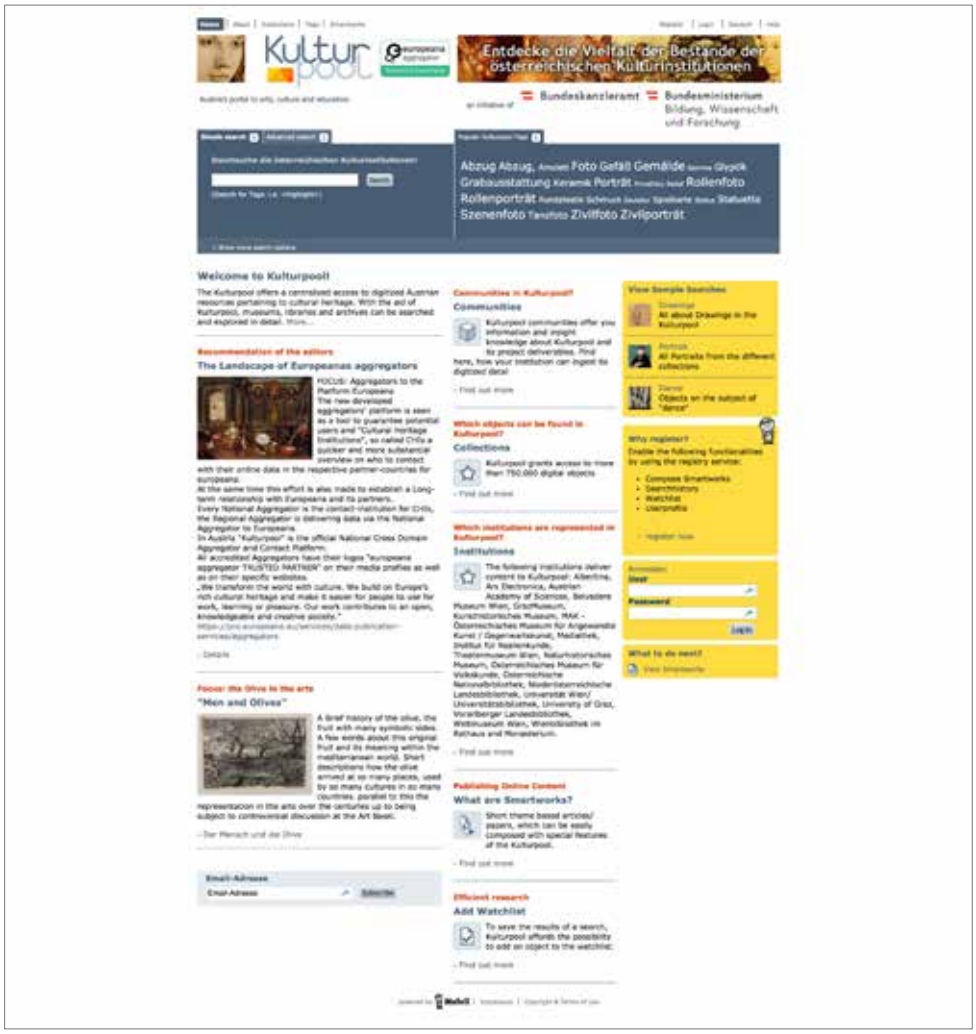


Fig. 10
Kultur Pool homepage.

Fig. 11
Kultur Pool layout analysis.

Aggregator: SearchCulture.
By: EKT aggregation.
Website: https://www.searchculture.gr/

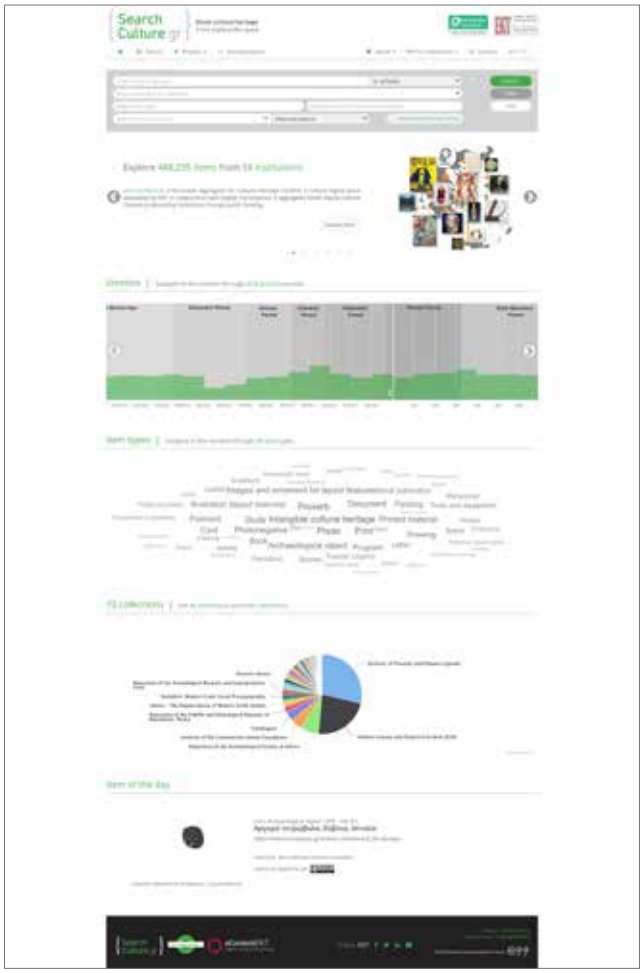


Fig. 12
SearchCulture homepage.

Fig. 13
SearchCulture layout analysis.

Aggregator: The European Library.
Funded by: Europeana Foundation.
Website: http://www.theeuropeanlibrary.org/

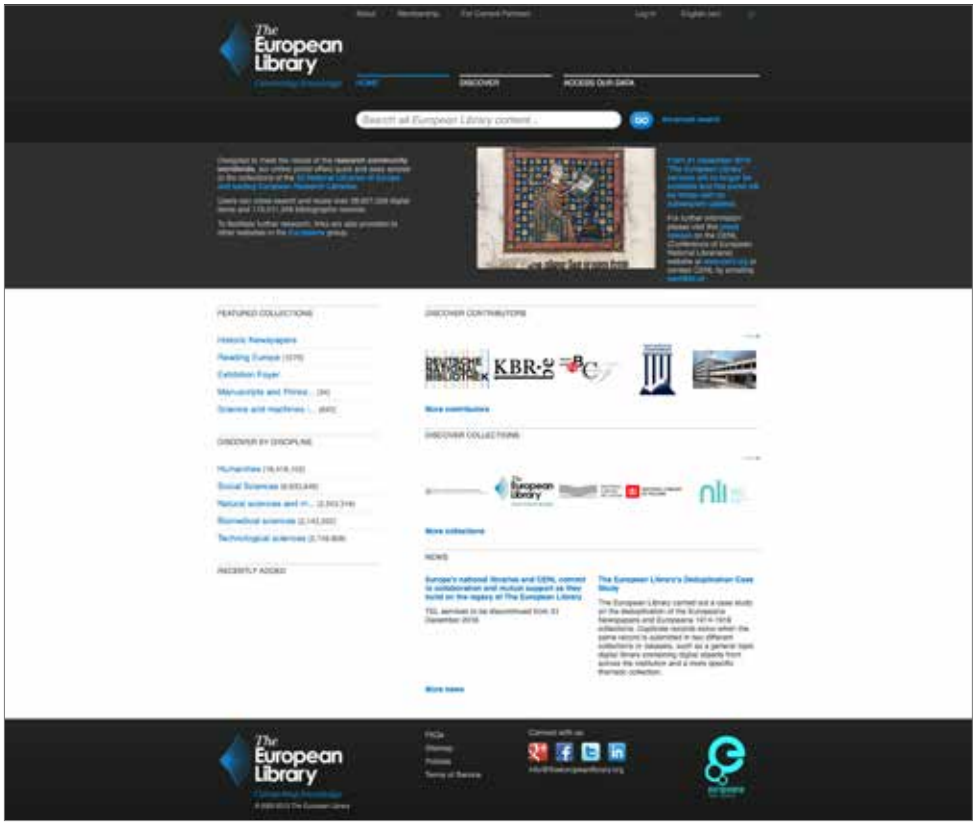


Fig. 14
The European Library homepage.

Fig. 15
The European Library layout analysis.

Aggregator: Wikimedia Commons.
Founded by: Wikimedia Foundation.
Website: https://commons.wikimedia.org/

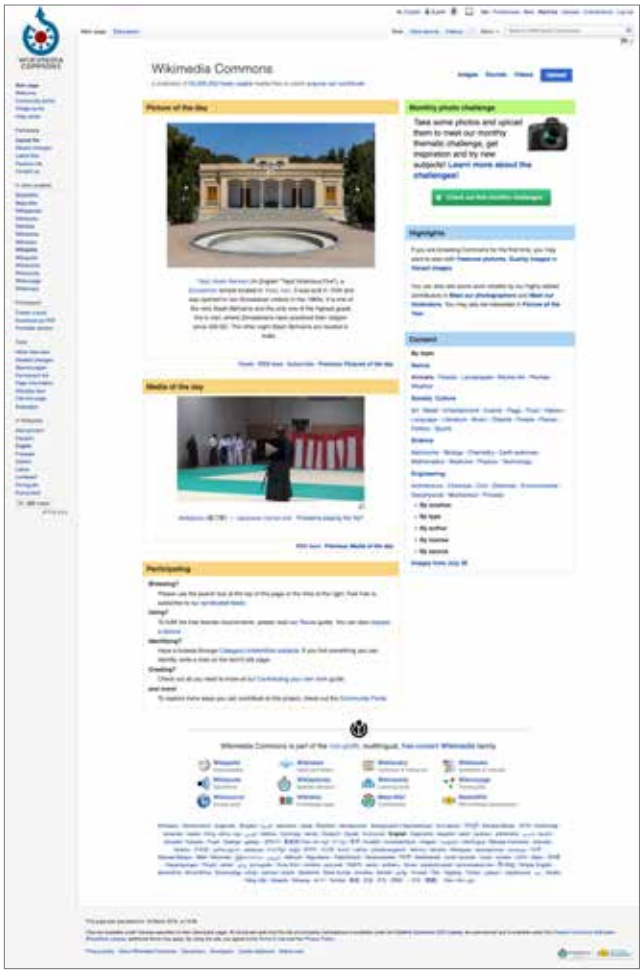


Fig. 16
Wikimedia Commons homepage.

Fig. 17
Wikimedia Commons layout analysis.

Quantitative comparison of the layouts

The bars show the percentage of area an element covers across the cultural content aggregators.

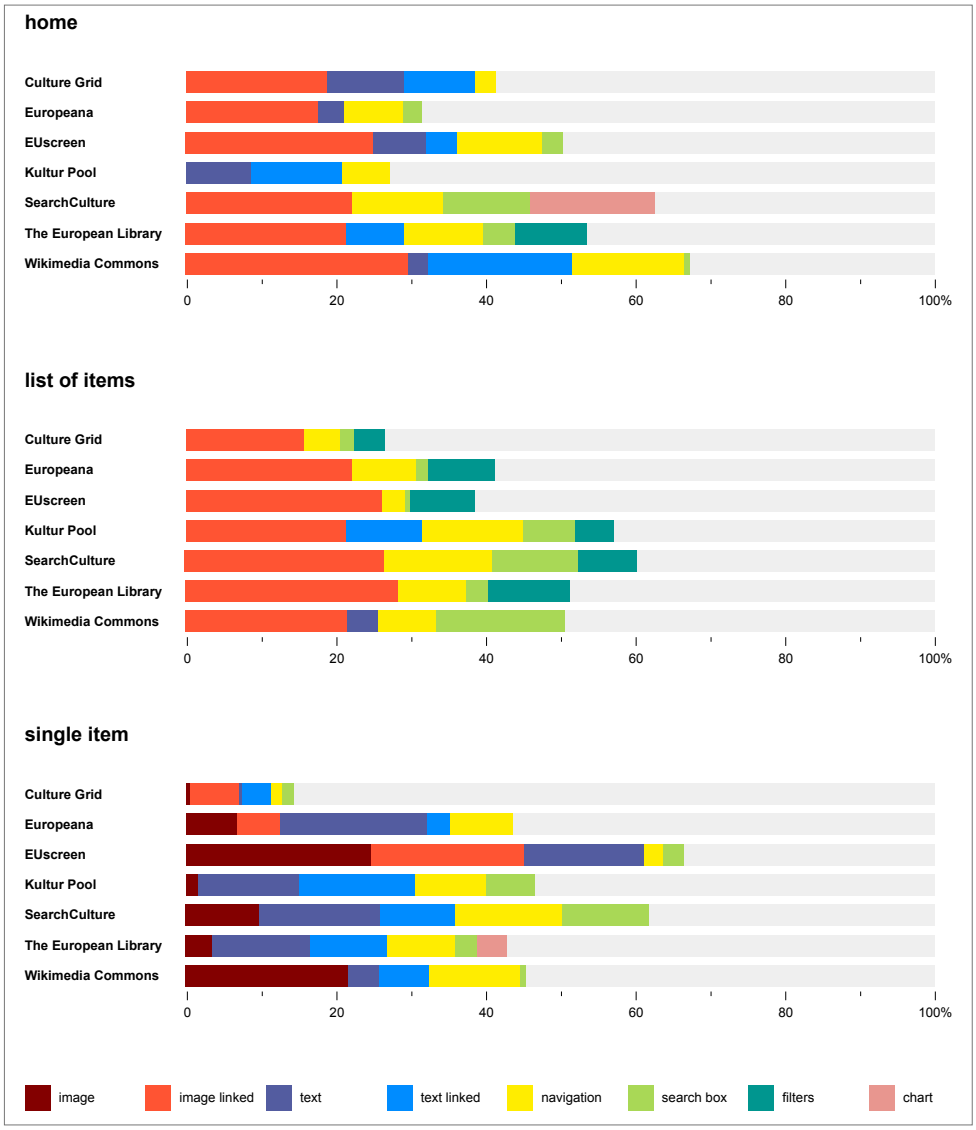


Fig. 18
Quantitative comparison of the layouts of the case studies.

Other relevant cultural content aggregators:

- Archives Portal Europe
  Website: https://www.archivesportaleurope.net/
- Deutsche Digitale Bibliothek
  Website: https://www.deutsche-digitale-bibliothek.de/
- EFG - The European Film Gateway
  Website: https://www.europeanfilmgateway.eu/
- Hispana
  Website: http://hispana.mcu.es/es/inicio/inicio.do
- Moteur Collections
  Website: http://www.culture.fr/Ressources/Moteur-Collections
- Swiss National Library
  Website: https://www.helveticarchives.ch
- The National Library of Finland
  Website: https://www.kansalliskirjasto.fi/en

Survey for end-users of cultural aggregators

Period: June - July 2019.

People who received the survey: 42.
People who replied: 18.
Composition:

- Web writers: 8
- Journalists: 4
- Bloggers: 4
- Wikipedians: 2

Questions
"I would be grateful if you can reply (even briefly) to the following questions":

1. How did you access the picture? (ex: official website, Google, Wikipedia etc.)
2. It was easy to find the image you were looking for? (It was well indexed/categorized/displayed on the online archive)
3. What kind of issues do you usually experience when you look for an image in an online digital archive? (ex: misleading/complex categorization, lack of related information, low-quality images etc.)

Way to obtain open digital surrogates (in descending order):

- Google search: 6;
  - Other content aggregator (mostly used for searching for low quality images): 5
  - Wikimedia Commons;
  - Imagno;
  - Pinterest;
- Digital archive of the GLAM: 4
- (Library of Congress, the Met Museum, the Yale University, "the excellent Lewis Walpole Library", RijkMuseum);
- Agencies: 3.

Cultural Content Aggregator (archive), Accessibility issues/strong points:

- Commons
  - "The way images are to be categorized" >Wikipedia/Commons Community
  - "The limited search options" > Wikipedia/Commons Community
  - "Sometimes, in Wiki Commons, the images are of low quality" > Bloggers
- Google Search
  - "Images easy to find, Yes pretty easy" > Bloggers
  - "It is very helpful in finding related images" > Bloggers
- Online archives (in general)
  - "Images might be descripted in a catalogue, but there is no visible image /no digital version/no scan available" > Bloggers "Images might be visible, but in not sufficient low-res quality"
  - > Bloggers "The question if there is a copyright or an open licence might not be made sufficiently clear, especially for users in a foreign language (legal terms can sometimes be puzzling and complicated, even in cases which are supposed to be "easy")"
  - > Bloggers

- "Navigation between the main menu, listings and the screen with the one single image might sometimes not be perfect"
  > Bloggers
- "The amount of related information for any item might differ and not be sufficient but, sometimes, this is explainable, in case the information is just missing. (I think sometimes it could be moreand even more complex!)
  > Bloggers
- "The related information might be quite incomplete or inaccurate (in some cases, you can get the impression that a careless "intern" had made the work and the supervisor did not double check).
  >Bloggers
- "In the British Library digital archives, the categorisation is sometimes misleading"
  > Bloggers
- "It contrasts with the British Museum (exorbitant) and the National Portrait Gallery (unbelievably complicated licensing arrangements - hopeless! They really are a nightmare to deal with). I am fortunate that most of the images are over 150 years old".
  > Bloggers
- "I strongly believe that all museums should make all their material available online without charge. The British Museum is my pet hate: it is funded by taxpayers - I have therefore already paid for their keep. The same applies to the Royal Collection - as a taxpayer I believe that it is wrong that I should pay for copies. Her Majesty was given those items on behalf of all of her subjects - they belong to us".
  > Bloggers
- "Images are often in low resolution, or lack caption information".
  > Bloggers

- Other cultural content aggregators
  - "We are used to use search tools with the correct keywords".
    > Communication agencies

Other issues revealed:

- "It is difficult to find the right images because most of them have not been digitised".
  > Bloggers
- "Need to categorize subjects differently because of time issues. I always search for help to minimize any time loss with searching".
  > Communication agencies

Survey on GLAM Culture Hub user interface

Period: July - August 2019

People who received the email: 46.
I estimated that, between social networks and the GLAM newsletter, around 200 people may have seen the survey page.
People who replied: 18.

Email sent
The email was translated and sent both to English and Italian speaking contacts.

Dear <name>,
I hope this email finds you well.

Following the research on the usage of open cultural content, I designed an interactive prototype of a cultural content aggregator. The aim is to identify interface characteristics that may foster the access and use of the digital heritage.

At the following link, you can find the interactive prototype of GLAM Culture Hub (the interface is not optimized for smartphones). It is based on static images with the addition of some sensible areas. Thus, functions cannot actually be performed.
https://gprofeta.invisionapp.com/public/share/UAWTM5SB8

I would be grateful if, after browsing the web platform, you can complete the following survey (it will take just 10 minutes to complete):
https://docs.google.com/forms/d/e/1FAIpQLSci0WdPMok9i3mu3ttf3agXS_Ii5D9aSthlDW8XUKF3l-Mi7A/viewform

The collected data is anonymous, but you can leave your contact information if you would like to stay updated on the follow up of the research.

Thank you in advance for your time.

Best regards,
Giovanni Profeta

Survey introduction
GLAM Culture Hub is an interactive prototype of a cultural content aggregator (such as Europeana and Wikimedia Commons) designed under the framework of the PhD thesis by Giovanni Profeta, focusing on the interface characteristics that may foster the access and use of digital images released under open licenses by Galleries, Libraries, Archives and Museums (GLAMs).

At the following link, you can find the interactive prototype of GLAM Culture Hub
https://gprofeta.invisionapp.com/public/share/UAWTM5SB8

In order to evaluate the characteristics of the visual interface, we kindly ask you to fill out the following survey (it will take 10 minutes to complete).

If you would like to know more information about the research project you can read the following page:
https://meta.wikimedia.org/wiki/Map_the_GLAM

1. Gender
one choice
a. Male
b. Female
c. Other
d. Prefer not to say

2. Age
one choice
a. <17
b. 18-24
c. 25-34
d. 35-44
e. 45-54
f. 55-64
g. 65 or more

3. Higher education
one choice
a. Less than a high school diploma
b. High school degree or equivalent
c. Bachelor's degree
d. Master's degree
e. Doctorate
f. Other

4. Profession
one choice
(Required)
a. Curator/archivist (working for a cultural institution)
b. Administrator of a digital archive (or a cultural content aggregator)
c. Designer
d. Writers/Journalist
e. Other
> please, specify.

5. Do you use open cultural content for professional, research or creative purposes?
one choice
(Required)
a. Never
b. Rarely
c. Sometimes
d. Frequently
e. Always

6. How do you evaluate the following features?
1: not useful, 2: slightly useful, 3: moderately useful, 4: very useful, 5: extremely useful
(Required)
a. The double chart to filter items (in home)
b. The search box
c. The related items
d. The map of the GLAMs
e. The list of the GLAMs
f. The user favorite items
g. The sharing options
h. The form for adding tags to a picture
i. The form to suggest an edit to a picture

7. Please respond to the following statements by selecting the option that best express your own opinion.
A: disagree, B: somewhat disagree, C: neither agree nor disagree, D: somewhat agree, E: strongly agree, F: I do not know
(Required)
a. Information is clear and well organised on the screens
b. The design throughout the system is consistent
c. The various functions in this system are well integrated
d. I think that the system is easy to use
e. I think that I would like to use this system

8. Rate each of the following aspects:
A: Strongly dislike, B: slightly dislike, C: neither dislike nor like, D: like a little, E: Love it, F: I do not know
(Required)
a. The general layout
b. The charts
c. The featured items
d. The textual filters
e. The overall interface

9. Which feature did you like the most? Why?
text area

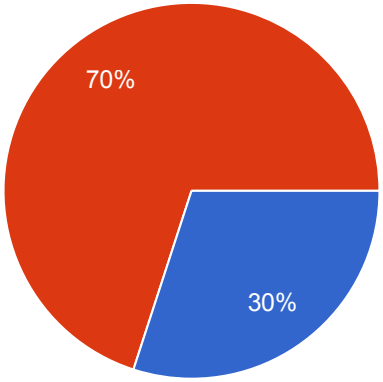
10. Which feature you did not like? Why?
text area

11. Other feedback/suggestions
text area

12. Contact information (if you want to stay updated about the project)
text area

Results

General information



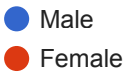
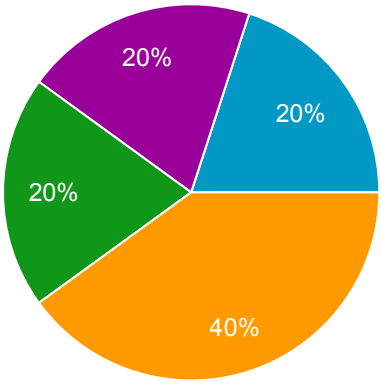


Fig. 19
Gender of the survey participants (on the top).



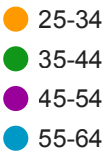


Fig. 20
Age of the survey participants.

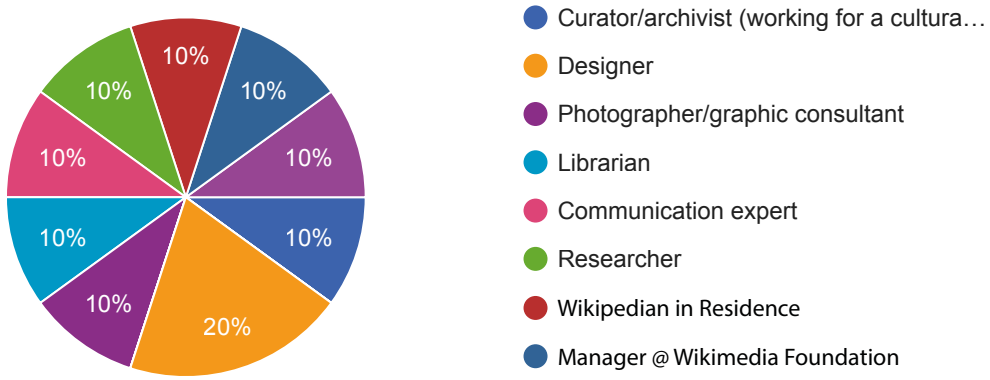


Fig. 21
Professions of the survey participants (on the top).

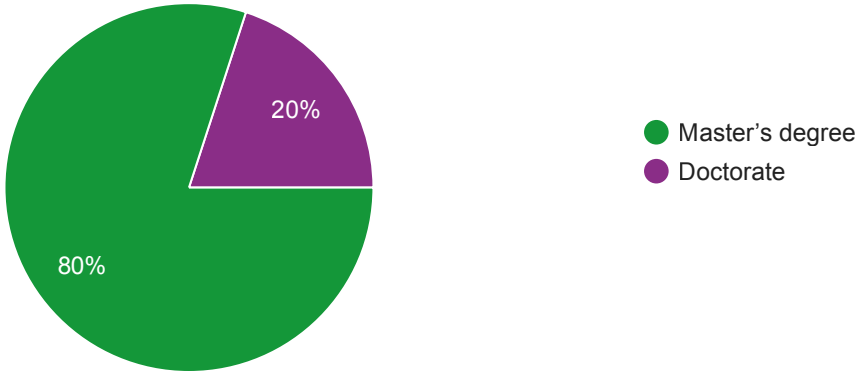


Fig. 22
Higher education of the survey participants.

Other aspects

Rate each of the following aspects

Which feature did you like the most? Why?
- I like the layout of all the pages and their colours.
- The thumbnails because of immediate understanding and with information that is well structured. All the functions (related item, sharing, tag) are extremely useful. Perhaps the edit is a bit unusual, but very sensible. I would reduce the dedicated box anyway.
- News - useful to discover other topics.
- Download and sharing options.

Which feature you did not like? Why?
- The characters of some texts are too small.
- I think the double chart (timeline and list of names), in the homepage, is a bit confusing. I would specify "choose by author" and "choose by period", or I would put the timeline and the list of authors side by side, so you can better understand the link, and display everything on the page as soon as I open (instead of having to scroll). The relationship is there, and it is visible with the use of colour, but in my opinion, it is not so intuitive. I also like related items, selection of images featured there seems to be pretty random.

Other feedback/suggestions:
- "Where is the map's pin legend? I don't understand the difference";
- "Many features are difficult to test because they are not working in my 2 browsers".

Related projects

During the Ph.D. studies, three related projects were made.

Wikipedia Primary School
2015—2017
Research team: Dr. Iolanda Pensa (principal investigator in Switzerland), Dr. Tobias Schönwetter (principal investigator in South Africa), Dr. Luca Botturi, Florence Devouard, Giancarlo Gianocca, Erica Litrenta, Giovanni Profeta, Marta Pucciarelli, Kelsey Wiens.
Project website: https://meta.wikimedia.org/wiki/Research:Wikipedia_Primary_School_SSAJRP_programme.

The research project was led by the Laboratory of visual culture at the University of Applied Sciences and Arts of Southern Switzerland (SUPSI) and the University of Cape Town. It was funded by Swiss National Science Foundation and South Africa National Research Foundation.
Wikipedia Primary School Project aims to evaluate methods to foster educators and experts to contribute to Wikipedia articles about topics related to the primary school curriculum in South Africa. The goal of the data visualization project is to visualise the state of the art of articles and their evolution over the project duration.

Visualizing knowledge gaps
2017
Research team: Dr. Serena Cangiano, Giovanni Profeta, Marco Lurati, Fabian Frei, Dr. Iolanda Pensa, Florence Devouard, Michele Mauri.
Project website: https://meta.wikimedia.org/wiki/Workshop_Wikidata_SUPSI.

The two four-day workshops were made within the framework of the "Wikipedia Primary School" project.
The workshop aimed to respond to the following questions: how to identify knowledge gaps on the Wikimedia projects? Which knowledge is missing on the wikimedia projects from Africa? Which knowledge is missing from the South African primary school curriculum (please refer to our selection of articles/topics)? The workshop relied on the idea that Wikidata is the best tool to trace what is available on the wikimedia projects.

GLAM visual tool
2016
Research team: Dr. Iolanda Pensa, Giovanni Profeta.
Project website: https://meta.wikimedia.org/wiki/Research:GLAM_visual_tool.

The research project was developed by the Laboratory of visual culture at the University of Applied Sciences and Arts of Southern Switzerland (SUPSI) and supported by Wikimedia CH.

The project aims to design a visual tool that GLAMs can use to visualise the impact of their collections within Wikipedia and its sister projects. The project involved the ETH-Library of Zurich and Synapta, a company focusing on data analysis and management based in Turin.