

Organizing subject access to cultural heritage in Swedish online museums

Swedish online
museums

Koraljka Golub and Pawel Michal Ziolkowski
Linnaeus University, Vaxjo, Sweden, and
Goran Zlodi

*Department of Communication and Information Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia*

211

Received 5 May 2021
Revised 14 September 2021
Accepted 16 September 2021

Abstract

Purpose – The study aims to paint a representative picture of the current state of search interfaces of Swedish online museum collections, focussing on search functionalities with particular reference to subject searching, as well as the use of controlled vocabularies, with the purpose of identifying which improvements of the search interfaces are needed to ensure high-quality information retrieval for the end user.

Design/methodology/approach – In the first step, a set of 21 search interface criteria was identified, based on related research and current standards in the domain of cultural heritage knowledge organization. Secondly, a complete set of Swedish museums that provide online access to their collections was identified, comprising nine cross-search services and 91 individual museums' websites. These 100 websites were each evaluated against the 21 criteria, between 1 July and 31 August 2020.

Findings – Although many standards and guidelines are in place to ensure quality-controlled subject indexing, which in turn support information retrieval of relevant resources (as individual or full search results), the study shows that they are not broadly implemented, resulting in information retrieval failures for the end user. The study also demonstrates a strong need for the implementation of controlled vocabularies in these museums.

Originality/value – This study is a rare piece of research which examines subject searching in online museums; the 21 search criteria and their use in the analysis of the complete set of online collections of a country represents a considerable and unique contribution to the fields of knowledge organization and information retrieval of cultural heritage. Its particular value lies in showing how the needs of end users, many of which are documented and reflected in international standards and guidelines, should be taken into account in designing search tools for these museums; especially so in subject searching, which is the most complex and yet the most common type of search. Much effort has been invested into digitizing cultural heritage collections, but access to them is hindered by poor search functionality. This study identifies which are the most important aspects to improve.

Keywords Digital cultural heritage, Online museums, Search interfaces, Subject searching, Controlled vocabularies, Information retrieval, Image retrieval

Paper type Research paper

1. Introduction

Ensuring online access to cultural heritage has been a key focus for many museums and cultural heritage institutions over the past few decades. Finding information objects online is

© Koraljka Golub, Pawel Michal Ziolkowski and Goran Zlodi. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Many thanks to Åsa Larsson of the Swedish National Heritage Board for providing key insights into some major aspects of the functionalities described in the Swedish heritage context. Special thanks to anonymous reviewers whose detailed suggestions helped greatly improve the paper.



directly dependent on the quality of search systems. In particular, searching by subject has proven to be very common amongst end users despite being the most challenging type of search due to the ambiguities of natural language. In order to help address this, subject indexing practices and standards (e.g. the [International Organization for Standardization, 1985](#)) prescribe assigned indexing in which subject index terms are taken from controlled vocabularies (such as thesauri or subject headings). These should also be applied in cross-search systems where subject searching is even more complex due to the increased heterogeneity of the collections. While international standards, policies and practices to support this are in place, the question is to what degree they have been followed in existing online services.

This study aims to investigate to what degree online web services providing access to cultural heritage from Swedish museums support subject searching and retrieval. To this end, a total of nine cross-search services and 91 Swedish museums' websites were evaluated in the period 1 July to 31 August 2020. The online services were evaluated against a set of 21 criteria.

The remainder of the paper is structured as follows. In the next section (2 Background), the rationale for the study is presented, describing user requirements and the established means to meet those requirements, referring to previous research and standards; this includes the special challenges in subject analysis involving museum objects. Sampling and methods are described in the third section (3 Methodology). The results are presented and discussed with regard to their implications for search, access and interoperability in [Section 4](#) (Results) and summarized along with guidelines for future research in the final section (5 Conclusion).

2. Background

2.1 Subject indexing and searching

Subject searching is common in online search systems such as library catalogues ([Hider and Liu, 2013](#); [Hunter, 1991](#); [Villén-Rueda et al., 2007](#)), online museums ([Baca, 2004](#); [Liew, 2004](#)), bibliographic databases ([Siegfried et al., 1993](#)), repositories ([Heery et al., 2006](#)), discovery services ([Meadow and Meadow, 2012](#)) and related digital search services ([Patel et al., 2005](#)). In comparison to known item searching (e.g. queries for information objects whose title, author, etc. is known beforehand), searching by subject is much more challenging. This is the result of difficulties in formulating search queries with insufficient knowledge of the subject matter at hand and/or insufficient knowledge of information searching (i.e. how to formulate a search query to reflect the information need), as well as challenges arising from semantic ambiguities inherent to natural language such as polysemy, homonymy and synonymy. Terminological polysemy leads to the retrieval of irrelevant results: in large databases, this may mean too many results to review manually. Synonymy presents challenges to effective searching by placing the burden on the searcher, who would ideally need to include all possible synonyms in a query in order to obtain a comprehensive set of results. Homonymy leads to queries that often end up producing false positives.

In order to alleviate these problems, online search services should use assigned indexing, a process in which subject terms are taken from established indexing systems such as subject headings systems, thesauri and classification systems. These are designed to help the user select a more specific concept to increase precision, a broader concept or related concepts to increase recall, to help the user disambiguate between homonyms, or to discover which term is best used to name a concept. In addition, hierarchical browsing of classification schemes and other systems with hierarchical structures could help the user improve their understanding of their information requirements and to formulate their queries more accurately.

The international ISO indexing standard of 1985, which was confirmed in 2020 ([International Organization for Standardization, 1985](#)), prescribes general techniques for subject indexing and clearly states that these are to be applied “by any agency in which human indexers analyse the subjects of documents and express these subjects in indexing terms” ([International Organization for Standardization, 1985](#), p. 1), defining documents to be “any item amenable to cataloguing or indexing, specifically including also non-print media and three-dimensional objects or realia”. The standard gives a document-oriented definition of manual subject indexing as a process involving three steps: (1) determining the subject content of a document; (2) a conceptual analysis to decide which aspects of the content should be represented; (3) translation of those concepts or aspects into a controlled vocabulary.

2.2 Special characteristics of subject indexing in museums

[Will \(1993\)](#) discusses how the principles of subject indexing museum objects are the same as for printed publications. Indeed, to some extent, museum professionals who document museum collections (custodians, documentalists, registrars) approach objects in the same way as they approach documents (document-like objects, [Caplan, 1995](#)) when conducting subject analysis and indexing, which allows them to share metadata standards and documentation methodologies with archives and libraries.

However, when conducting subject analysis and indexing of museum objects in the broad domain of cultural heritage, museum professionals need to take into account some special characteristics of those objects, of which the key characteristics are described below.

2.2.1 Heterogeneity of museum object types and broad cultural heritage. The International Council of Museums ([ICOM Statutes, 2007](#), p. 2) defines a museum as “... a non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment”. Compared to archives and libraries, museums collect a vast number of heterogeneous object types. Also, when we speak about museum objects, we predominantly mean unique objects: well over 99% of all museum objects are unique objects/works, entities that do not come as multiple copies or examples of manifestation of some work. With respect to their subject, most museum objects do not even have narrative content, but are instead designated by their material, form and function. Therefore, in subject analysis and indexing, it is important to identify and represent not only motives depicted in an object (*ofness*) or what an object/work is about (*aboutness*) but also what an object is *per se* (*isness*) and what its function is.

2.2.2 *Isness, aboutness and ofness.* To further elaborate the concepts of *isness*, *ofness* and *aboutness*, let us take *Hagia Sophia* (Istanbul) as an example in two different cataloguing procedures:

- (1) When cataloguing a photograph (the Object/work type is *photograph*) depicting *Hagia Sophia* from 2020, the title should be “Hagia Sophia” representing work (*ofness*), and one of the terms of the subject index should be “mosque” (*aboutness*).
- (2) When cataloguing the *Hagia Sophia* itself as a built work, according to the book *Cataloguing Cultural Objects: a Guide to Describing Cultural Works and Their Images (CCO)*, three values should be recorded in the Object/work type element ([Baca et al., 2006](#), p. 57): “cathedral”, “mosque” and “museum”; the general subject term should be “architecture”; and the Sspecific subject terms should be “cathedral”, “mosque” and “museum” (*isness*), that is, the same as the Oobject/work type.

The function of the Hagia Sophia building has changed over time: It was built in 537 as Byzantine Christian cathedral; in the 13th century it became the city’s Roman Catholic

cathedral and in 1453, after the fall of Constantinople to the Ottoman Empire, it was converted into a mosque. In 1935, the Republic of Turkey established it as a museum; in 2020, it was reopened as a mosque. Consequently, when cataloguing a building *per se* (example 2) we must index all the functions that the building had, and when cataloguing a photograph showing Hagia Sophia (example 1), then it is crucial to identify the time when the photograph was taken and index only the function the building had at that very moment.

The short definition of *isness* from the Categories for the Description of Works of Art (CDWA) guidelines is based precisely on the overlap of object/work type and subject index elements: “If the work *is* the subject term, this is called *isness*” (Baca and Harpring, 2016). *Isness* denotes what a work is or which class of objects it belongs to; therefore, these kinds of subjects are based on the form or type of the object or its genre.

Values for *isness* are ideally taken from controlled vocabularies such as the Art and Architecture Thesaurus’ (AAT) Objects Facet (i.e. visual works by form: medals, tapestries, diptychs). In addition to the object/work type, the function of the object and changes in its function over time are essential for documenting utilitarian objects and various forms of industrial and technical heritage. In this respect, the CDWA clearly recommends the following: “Works that have a primarily functional purpose, such as architecture and utilitarian objects, should also be analyzed for subject, including the work’s function and/or form” (Baca and Harpring, 2016). AAT’s Objects Facet commonly lists concepts/object names (e.g. parade armours, wedding dresses) in various classes that provide guide terms related to function (e.g. <armours by function>, <dresses by function>).

In practice, however, when creating metadata using established standards, the *isness* subject type is problematic because it overlaps with the Object/work type metadata element. The term “pocket watch” must be recorded in the Object/work type element in CDWA (or in the semantically equivalent Object name element in SPECTRUM). Moreover, *isness* overlaps with elements in the group Object history and association information in the SPECTRUM standard: in this group, in addition to the element in which the function is recorded, there are elements that determine the spatial and temporal context of functions. However, should we record this term as a subject as well? Is that unnecessarily redundant, or would this redundancy help users search and retrieve subjects? This problem requires further research and feedback from user studies.

While on the one hand *isness* is related to a work’s function or form, and is therefore used mostly for utilitarian objects and architecture, *ofness* and *aboutness*, on the other hand, are most often encountered in the field of fine arts. *Ofness* is about what a work depicts, what a non-expert viewer could see and recognize in the visual content of the object (e.g. plants, animals, objects), and *aboutness* is related to narrative, thematic, iconographical or symbolic meaning. *Ofness* and *aboutness* reflect the dual nature of the visual symbol, which further contributes to the complexity of image analysis (Jack, 2001): an image represents both a physical object (e.g. owl) and an idea (e.g. wisdom). Both of these aspects of the object are important for subject access. For example, one user may want to find works that simply represent owls (*ofness*). As owls commonly represent wisdom in Western civilization; another user might be interested in wisdom as a subject (*aboutness*). At the *aboutness* level of analysis, metadata creators should be especially careful of culturally conditioned symbolisms, meanings and interpretations, as in this example: in most Native American legends, owls are a symbol of death.

Furthermore, Erwin Panofsky distinguishes three levels of the analysis of subject matter or meaning that are applicable to artistic and utilitarian objects that have a certain visual content (e.g. pictures, drawings, tools with depicted content) or to objects whose material is designed to represent a certain form and content (e.g. sculptures, totem poles, origami, lace):

- (1) At the first level, by which we describe the *primary or natural subject matter*, we determine the *factual* and *expressional* visual content. All these objects and features form a list of artistic motifs, and their enumeration about a particular work of art constitutes a pre-iconographic description (i.e. *ofness*).
- (2) At the *secondary or conventional subject matter* level, we connect the presented artistic motifs with certain concepts, themes or ideas, or stories and allegories, and thus we discover the secondary or conventional meaning (i.e. *aboutness*).
- (3) *Intrinsic meaning or content* is achieved by interpreting motifs, themes and allegories from previous levels of analysis and constituting the world of symbolic values. It is the subject of iconography in a deeper sense, which implies a method of interpretation that is more of a synthesis than an analysis (Panofsky, 1993, pp. 53–55).

To provide a distinction between the aforementioned levels of subject analysis, indexing and interpretation, CDWA metadata categories include the General Subject Type and Specific Subject Type elements with a controlled set of values limited to *description, identification, interpretation, isness, aboutness, ofness*.

2.2.3 The problem of translatability between different media on subject indexing. Svenonius (1994) pointed out the key challenges of how to represent in subject index terms a topic from an object represented in media that does not use any words, but is instead represented in visual (art) or aural (music) terms. In comparison, when conducting subject analysis and indexing of textual documents, we always use the medium of language: while examining documents, we read natural language; while determining their subjects, we use language concepts; and while selecting index terms, we use various artificial languages. But museum and heritage objects (with relatively few exceptions) typically do not have content encoded in text that we could read. Most of them do not have narrative content; and in those which have, this content is encoded primarily in a visual system that needs to be determined, expressed into natural language concepts and finally translated into metadata that will represent the work in an information system.

2.2.4 The importance of visual representations. For all of the reasons mentioned above, the existence of high-quality visual representations is crucial for museum and heritage resources. As access to image-based resources is fundamental to research, scholarship and the communication of cultural knowledge, it is no longer enough to provide only simple (thumbnail or full screen) images, but it is necessary to enable functionalities such as deep zooming or annotation capabilities that enable transfer of knowledge about the various characteristics of heritage objects mentioned so far.

In that context, the International Image Interoperability Framework (IIIF – [International Image Interoperability Framework, 2021](#)) represents an important standardization initiative which defines, develops, cultivates and documents methods for access and use of high quality image resources, not only deep zoom and annotation capabilities but also features for image comparison and manipulation (rotating, setting brightness, contrast, etc.) as well as a set of common application programming interfaces (APIs) that support interoperability between image repositories (IIIF Image API and IIIF Presentation API). An example of such interoperability is enabling the display of an image from one repository in another information context without the need to create a copy of the image.

2.2.5 Content-based image retrieval and computer vision in subject indexing. While in the context of text-based documents we talk about full-text search, in the context of visual resources we talk about content-based image retrieval (CBIR), a methodology for retrieving images from databases of images based on visual content. In CBIR technology, content most often refers to various low-level visual features such as textures, colours and shapes that can be detected within the image and are then used as index terms. At the beginning of image

retrieval, a user expresses his or her imaginary intention into some concrete visual query which can be *query by example image*, *query by sketch map*, *query by colour map*, *query by context map*, etc. (Zhou et al., 2017, p. 2). An example from Europeana, illustrated in Figure 1, shows the combination of textual search criteria (“painting”), the providing country (Sweden), the licence type (free reuse) and a query by colour (red and green).

Zhou et al. (2017) emphasize the role of deep learning (DL), belonging to a broader family of machine learning (ML) methods, as a part of the field of artificial intelligence (AI) in CBIR, which enables learning of high-level abstractions close to human cognition processing. This could help museums not only to recognize real-world objects (*ofness*) automatically and identify particular people with facial recognition systems but also to obtain high-level understanding of images (e.g. ambience, moods, expressions).

2.3 Metadata guidelines and standards related to subject indexing

2.3.1 *Conceptual models.* In the museum’s community, the standard corresponding to the Library Reference Model (LRM) developed by IFLA (International Federation of Library Associations and Institutions) is the ISO standard CIDOC-CRM developed by the International Council of Museums (ICOM)’s International Committee for Documentation (CIDOC Conceptual Reference Model, ISO, 2006). This is described further in the section “User requirements” below. However, the two are rather different in that CIDOC-CRM focuses on events and processes while IFLA LRM primarily models the outcomes of the processes. In order to align the two models, the International Working Group on Functional Requirements

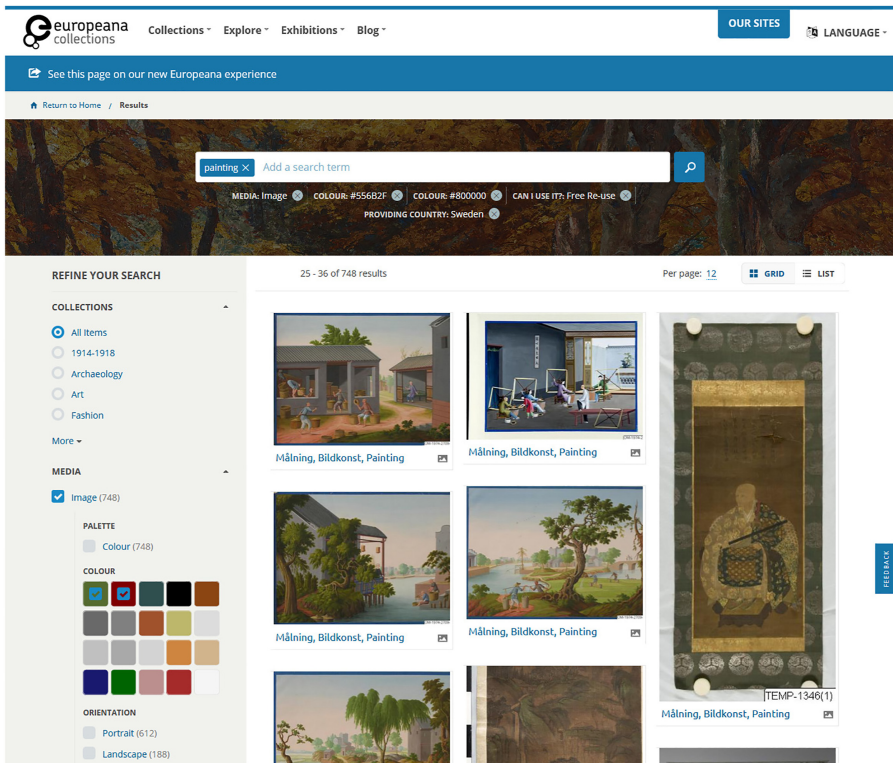


Figure 1.
Example from
Europeana portal on
searching by colour

for Bibliographic Records (FRBR)/CIDOC-CRM Harmonisation was established. It published FRBRoo (FRBR Object Oriented, [ICOM and CIDOC, 2006](#)) to serve as an ontology to facilitate the integration and interchange of bibliographic and museum information.

While CIDOC-CRM supports *aboutness* ([ICOM and CIDOC, 2002](#)) at a very general level, and attempts exist to evolve this further (see, for example, [Carboni and Luca, 2017](#)), we would like to point to how subjects are modelled in the IFLA LRM model, hoping that the FRBR/CIDOC CRM Harmonisation Group will take the FRBRoo development further towards modelling *aboutness* more specifically.

2.3.2 Subjects in data structure standards. Data standards for museum resources provide a good structure for recording subject-related information, with small differences that we hope will be harmonized in the future. The key reference point for the documentation and management of all types of museum collections is SPECTRUM, the United Kingdom's collection management standard that is increasingly adopted in other countries. SPECTRUM standardizes collection management procedures and data structure. Regarding subject indexing, one of the prescribed primary procedures of SPECTRUM Version 5.0, *Cataloguing*, defines the *Object description information group* that contains the *Content and subject information* subgroup. The CDWA comprises a set of guidelines for best practice in cataloguing and describing works of art, architecture, other material culture, groups and collections of works, and related images. CDWA is arranged in a conceptual framework that may be used for designing data models and databases, as well as for accessing information. CDWA includes around 540 categories and subcategories of information ([Baca and Harpring, 2016](#)). CDWA guidelines define a group of metadata categories called Subject matter which contains 17 subcategories related to the subject.

In order to establish semantic interoperability between different metadata schemes, a number of mappings have already been implemented, such as the Metadata Standards Crosswalk included in the CDWA standard. In establishing interoperability in the museum community, the LIDO (Lightweight Information Describing Objects) XML metadata exchange schema ([Coburn et al., 2010](#)) plays a particularly important role. LIDO takes into account both SPECTRUM and CDWA data structures, and its development (currently in beta version 1.1) has been prompted by the need for better connectivity with LOD vocabularies and more functional data display in public catalogue interfaces.

2.3.3 Controlled vocabularies related to the subject. Here we would like to focus on a few key international classification systems and thesauri that serve as data value standards and provide vocabulary control for subject terms in museums: Iconclass, the Getty Vocabularies, and the Social History and Industrial Classification.

Iconclass is a classification system designed for art and iconography. It is the most widely accepted scientific tool for the description of subjects represented in images ([Iconclass, 2021](#)). The Iconclass classification is accessible through the web-based Iconclass Browser and is also available as linked open data (LOD) that enable interoperability in terms of linking, exchange and enrichment of metadata.

The Getty Vocabularies include the following series of highly influential multilingual thesauri and databases that are constantly evolving and being updated through a number of international initiatives and projects. These can also be used to control vocabulary in subject indexing. The AAT contains concepts and terms on objects, materials, techniques, styles, periods and other concepts related to art, architecture and the broader field of cultural heritage. The Cultural Objects Name Authority (CONA) is a relatively new terminology resource that compiles a list of important works of art and buildings. Examples of names in CONA are *Chayasomesvara Temple*, *Mona Lisa*, *Livre de la Chasse*, *Hagia Sofia* ([Harpring, 2018](#), p. 3). CONA could be used in cases when a work depicts another work. The Getty Iconography Authority (IA) is a relatively new terminology that focuses on subjects and topics (e.g. *Bouddha couché*, *Adoration of the Magi*) ([Harpring, 2018](#), p. 2). The Union List of

Artist Names (ULAN) is also important for recording names as subjects; it not only contains the names of artists but also the persons associated with them (teachers, families, collaborators) and the names of corporate bodies (museums, workshops, art groups). The Getty Thesaurus of Geographic Names (TGN) is important for recording geographical names as subjects. All Getty Vocabularies have a SPARQL endpoint and are available as LOD, which is important for linking and enrichment from other vocabularies (i.e. VIAF for people as subjects, GeoNames for places as subjects, etc.).

The Social History and Industrial Classification (SHIC) is used in social and local history museums as a means to make links between objects through their context and background (e.g. functions of objects, human activities). The classification system is divided into four top facets (community life, domestic and family life, personal life, working life), which together represent all aspects of human activity ([Social History and Industrial Classification, 2021](#)).

2.3.4 Data content and cataloguing rules and guidelines. There are only a small number of cataloguing rules in the museum community. The book *Cataloguing Cultural Objects: a Guide to Describing Cultural Works and Their Images (CCO)* is the most influential set of guidelines for cataloguing cultural works and their visual surrogates. It addresses subject indexing in more detail, published in 2003 and available in a digital open access edition from 2006 ([Baca et al., 2006](#)).

The next important data content standard to consider is CDWA, which largely overlaps with CCO in content and structure. More importantly, guidelines are further developed and upgraded (the latest version of CDWA was revised in 2016 by editors Patricia Harpring and Murtha Baca). Cataloguing rules for subject analysis and indexing in CDWA are described in detail in the *Subject Matter* information group, which, according to the guidelines, is also one of the mandatory elements for cataloguing works.

That said about standards catering for subject indexing in museums, the software interfaces seem to indicate that this practice in museums is underused. There are few empirical studies that analyse subject indexing in museums. A rare example is an OCLC study survey of core fields (agreed by the CDWA and CCO guidelines) from 2010 which shows that nine prominent American art museums apply 17 core fields more or less consistently, but that *subjectTerm* is a conspicuous outlier in that it is used by only two institutions ([Waibel et al., 2010](#)). Since then CIDOC, Getty, the Collection Trust and others are working hard to develop standards to improve this situation.

2.4 User requirements

Guidelines for providing subject access in information systems have their origins in cataloguing standards used in libraries and related information services. The objectives of library catalogues for subject access are anchored in Charles Ammi Cutter's "objects", as he called them, the purposes of which are to (1) enable finding an item of which the subject is known; (2) show what the library has on a given subject and (3) assist in the choice of a book based on its topical character ([Cutter, 1876](#), p. 5). These objectives have been an integral part of library cataloguing codes for nearly 150 years and continue to be so in the contemporary FRBR family of conceptual models for catalogue functionality. These are FRBR; Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD). They were consolidated into the IFLA LRM ([Riva et al., 2017](#)).

The IFLA consolidated model prescribes five user tasks, which then need to be translated into cataloguing rules to account for relationships between works, expressions, manifestations and items, as well as for relationships between topics and these works, expressions, manifestations and items. In the context of subject access, the IFLA LRM and FRSAD ([Zeng et al., 2011](#)) tasks of finding, identifying, selecting, obtaining and exploring could be defined as

- (1) Find: to find resources embodying works that are described by a given subject label, for example, search using a nomen that is used in a subject headings system or a classification scheme.
- (2) Identify: to clearly understand the nature of the resources found and to distinguish between similar resources, for example, those that are indexed by homonyms or those with the same topic but from a different perspective (e.g. different branches of a classification system like a virus from a zoological perspective vs a medical perspective).
- (3) Select: to determine the suitability of the resources found and to choose (by accepting or by rejecting) specific resources that seem the most relevant, for example, because of certain aspects, facets or approach of the subject described.
- (4) Obtain: to access the content of the resource.
- (5) Explore: to use the subject relationships between one resource and another to place them in a context, for example, to browse around related topics by using related terms in a thesaurus or similar; or to see narrower and broader terms or classes in order to understand the relationships between various nomens for an entity. Examples include the following: examine the variant names for a subject within a controlled vocabulary, survey the variant terms used in different contexts of use, which may include different languages; explore correlations between nomens for the same entity in different controlled vocabularies, for example, finding a thesaurus descriptor which corresponds to a classification number.

However, in neither libraries nor museums are these models of cataloguing standards put into practice with regard to subjects. While previous library cataloguing codes such as AACR2 (Anglo-American Cataloging Rules) did not mention subject cataloguing, the most recent cataloguing principles, Resource Description and Access (RDA), make an effort to point out that a subject representation or relationship to the subject of a work is needed (Kuhagen, 2015, p. 3; RDA Co-Publishers, 2017). However, this has not been extensively elaborated, so concrete guidelines for the practice of subject indexing are lacking.

In all cultural heritage institutions, the transition to the online environment has created the opportunity to engage a range of potential audiences. “While printed permanent collection catalogues are typically aimed at an exclusively scholarly audience, the Internet allows museums to engage multiple audiences simultaneously . . .” (The Getty Foundation, 2012). Walsh, Clough and Foster (2016) in their literature review of archives, libraries and museums identify user groups coming to online services divided by three sets of criteria: (1) the level of subject knowledge, identifying experts, semi-experts (including hobbyists) and non-experts; (2) information need, distinguishing between general visitors (interested in opening hours, cost, etc.), educational visitors (interested in more detailed information to plan a visit and in project-based information) and specialist visitors (interested in the museum’s collection which they explore online); and (3) motivation and role, identifying explorers (driven by personal curiosity), facilitators (engaging in a social experience with someone who they care about), experience seekers (looking for things and ideas that are intellectually important in the community), professionals/hobbyists (satisfying the information need of a specific subject matter) and rechargers (wanting to emotionally and intellectually recharge by viewing art). Other researchers have identified additional user motivations: users visit online museums to conduct research which can be of personal, student or professional interest (Fantoni *et al.*, 2012; Villaespesa *et al.*, 2015; Villaespesa, 2017), to find inspiration, for enjoyment, to access art news (Villaespesa *et al.*, 2015), for casual browsing (Fantoni *et al.*, 2012), to pass time (Skov and Ingwersen, 2014; Walsh *et al.*, 2018) or to plan their visit (Fantoni *et al.*, 2012; Skov and

Ingwersen, 2014; Villaespesa *et al.*, 2015). While this may hold true in general, the specific ratios of the above motivations will differ from one museum to the other. For example, planning visits to the museum was the dominant motivation at the Tate, UK, at 41% (Villaespesa *et al.*, 2015), less common at the Copenhagen Military Museum at 21.5% (Skov and Ingwersen, 2014), and at only 8.1% at the Metropolitan Museum of Art (Villaespesa, 2017).

With respect to search preferences, a comprehensive study of seven museums of diverse collection types belonging to the National Museums Liverpool (NML) digital collection (Walsh *et al.*, 2018) shows that the majority (67% of the general public and 59% of non-professional researchers) prefer browsing to a search box. Different results were obtained for the National Museum of Military History in Copenhagen, Denmark (Skov and Ingwersen, 2014), where users chose browsing only in exploratory searching while preferring free text search in all other types information search. The study also showed that the majority of advanced users find most of the descriptive metadata elements useful in their searches but did not give further detail about which elements are used or any detailed analysis of the usefulness of the subject element, if used at all.

Related research has shown that while some online museums support subject access (Liew, 2004), Trant *et al.* (2006b) state that others (especially art museums) do not because providing subject access is not necessary for its operation (unlike object registration, inventory, location control, etc.). In fact, the overall impression seems to be that many museums describe their collections in far too simple terms which includes the title of a work, the creator's name, dimensions and sometimes a picture of the museum's object (Fortier and Ménard, 2018); this in spite of the standards and guidelines outlined above (ISO, 1985; Baca *et al.*, 2006).

While a general overview of functional requirements for digital museum search interfaces is lacking, in part due to differences between museum types, an example by the National Gallery of Art lists a total of 75 functional requirements (The Getty Foundation, 2012), of which 31 requirements are related to search/retrieval. It is unclear from the report to what degree user requirements were taken into account in the research from which the list was derived, but we nonetheless keep it as a key reference, as a rare example of what is recommended. The requirements include field-based keyword search, auto suggestion of available terms, Boolean operators, refinement of search results by modifying the search criteria, preserving search history and allowing combination/modification of earlier search/browse sessions with the option to add and subtract browse/search facets into the current or past browser search result; support for search/browse functionality with a synonym ring, authority files and provision of alternatives to those entered by the searcher; expansion of results with broader terms; faceted browse searching on criteria which include onness, aboutness, tag clouds, object type, etc.; linked terms from search results to other results linked to the same terms; controlled vocabularies including at least ULAN, AAT, TGN, ICONCLASS; highlighting keywords from the search phrase in the results. Other relevant functionalities include providing contextual help to users (display of a pop-up short description upon hovering over a function) as well as display of a visual timeline of artists and works of art and other world events; visualization of artists, artworks and world events on maps (GIS – geographic information systems). In addition, a number of functionalities related to display, ranking and navigation of search results are listed.

Full-text searching is not enough. Knapp *et al.* (1998) established that the most effective way of online searching databases in the humanities is to combine free-text searching with the use of controlled-vocabulary indexing. Controlled vocabularies are particularly needed in large databases covering many subjects (Markey, 2007; Tibbo, 1994) as well as in databases of primary sources (Bair and Carlson, 2008) such as museum objects, which cannot be queried using full-text searches alone. Tibbo (1994) makes the point that the exponentially increasing

volume of information objects available online leads to information overload and entropy, rather than increasing benefit from access to information. Although full-text indexing works for some tasks, for others it creates information overload and prevents the searcher from gaining a comprehensive overview on a topic: if a query returns thousands of retrieved documents, few searchers will browse beyond the first dozen or two hits.

To counter high recall with hundreds or thousands of hits and low precision, specific subject indexing should be implemented, involving (1) indexing policies that promote a high level of specificity and (2) indexing languages that are deep and detailed for any given topic, especially for large databases and cross-search services with tens of millions of records. The indexing language needs to be extensive in order to account for the fact that any topic can appear in many different contexts, and topics may be addressed from a very wide range of different perspectives. Furthermore, specific disciplines will require their own specific indexing languages, rather than a one-size-fits-all approach (Tibbo, 1994).

From research in library information systems we know that researchers from different disciplines have varied needs when it comes to subject access in bibliographic databases (for an overview, see Golub *et al.*, 2020), showing that in humanities research, ordinary users of cultural heritage primary sources in museums need a faceted approach to controlled vocabularies such as the Arts and Architecture Thesaurus for visual arts, rather than for pre-coordinated ones like subject headings. Faceted vocabularies are more suitable since they support high specificity and can account for the different facets that are important to humanities scholars, such as geographical, chronological and disciplinary terms (see Bates, 1996; Tibbo, 1994). Facet selection and query expansion based on such controlled vocabularies also need to be implemented effectively in search interfaces, which currently seem to be limited to experimental interfaces (e.g. Alani *et al.*, 2000; Tudhope *et al.*, 2006) rather than applied in practice across online search systems.

Furthermore, online cultural heritage used to be described mostly for museums' internal use or for subject specialists such as art historians (Trant *et al.*, 2006a). Enabling subject access, which is more useful for the general public, as it encourages exploratory searching, is widely acknowledged to be expensive and difficult. This gave rise to social tagging in museum interfaces and projects such as *steve.museum* (Trant *et al.*, 2006b). Social tags, in spite of being broadly criticized for their subjective nature, the lack of term consistency and other features of high-quality knowledge organization systems (KOS) (see, e.g. Trant, 2009a; Srinivasan *et al.*, 2009) hold the potential to fill the existing subject access gap, thus improving access to museum collections for the general public (Trant *et al.*, 2006b). Social tagging augments the professional description of art museums' collections: 77–85% of tags found in the research (the ratio depends on the kinds of objects and their specific context) are new terms not existing in the KOS in use; most of these are subject terms (39%) and genre terms (27%) (Trant, 2009b).

Future user studies should be undertaken to clarify user needs in the context of subject access to various types of museum collections. Investigation of *isness* type of subject is particularly important because the vast majority of museum objects (especially those from the archaeological, ethnographic, technical, not to mention natural history collections – all of which are very numerous compared to art collections) have no *ofness* or *aboutness* types of subjects at all. Rare exceptions are objects from the aforementioned collections on which there is a certain visual motif.

2.5 Cross-search services

In cross-search services, the most common issues affecting subject searching today are the inconsistency and incompleteness of metadata and the blending of controlled vocabularies, free keywords and full-text indexing (Dempsey, 2012; Fagan, 2011; Golub, 2016). Interoperability has been acknowledged as a key issue in cultural heritage contexts

(Koutsomitropoulos *et al.*, 2012; Seadle, 2010). A large number of national and international infrastructure projects are working on making cultural heritage collections interoperable with each other. Semantic Web standards and interoperability opportunities for cross-institutional searching and linking of cultural heritage data have been available for some time now, and many institutions today provide metadata and/or digital information objects to portals such as Europeana and World Digital Library that allow cross-searching of dispersed collections.

Europeana is a prominent cross-search service which combines metadata from thousands of libraries, archives and museums. The objects are described using different metadata standards, languages and indexing policies. To address the problem, Europeana developed a data model EDM (Europeana Data Model) based on 15 elements of the Dublin Core Metadata Standard, enriched with an additional 13 elements. Two of the metadata elements are subject related: *dc:subject* [the subject of the Cultural Heritage Object (CHO)] and *dc:type* (the nature or genre of the CHO) (Europeana Foundation, 2017, 2021). This has been identified as insufficient since it leads to inaccurate search results with high recall and low precision (Gaona-García *et al.*, 2017; Dobрева and Chowdhury, 2010). More recently, Europeana decided to adapt EDM to Schema.org (Freire *et al.*, 2020) because it is supported by major Internet search engines (Wallis *et al.*, 2017). The model uses semantic description languages using resource description framework (RDF) and simple knowledge organization system (SKOS) and links resources according to LOD principles (Gaona-García *et al.*, 2017).

Work on standards like CIDOC-CRM and FRBRoo is meant to enable sharing of metadata across institutions, with the idea of creating a one-stop shop for all potentially relevant resources. Europeana is perhaps the most comprehensive example of this idea coming to fruition. It is therefore especially important that an FRBRoo-EDM application profile is developed (Doerr *et al.*, 2013). Swedish museums and certain other cultural heritage institutions make their resources searchable in Europeana while also making them available via the Swedish cross-search service Kringla <http://www.kringla.nu/> (Swedish National Heritage Board, 2021).

2.6 Desirable features of online subject access

In summary, as we witness developments in digital scholarship, it is important to provide quality subject access to a vast range of heterogeneous information objects in digital services. This includes both primary and secondary sources. The general objective of subject indexing should be that it allows the user to find anything and everything in the collection (including cross-search collections) that is relevant to a certain topic, and this requires that controlled vocabularies need to be applied to ensure high precision and recall. In addition, calls should be made for highly specific subject indexing, the application of controlled vocabularies that are faceted rather than pre-coordinated to cater for a range of possible topics discussed from different perspectives, and the inclusion of named individuals as well as facets of space and time.

As a summary of the literature above and as discussed by Golub (2003, 2018), subject access in online information retrieval systems should involve the following first 18 options shown below. While this research focuses on online catalogues of Swedish museums, it addresses organizing subject access to broader cultural heritage. Despite special characteristics of museums, we believe these features need to be common on search interfaces across libraries, archives and museums. Here we add three additional image-related features that are important in museums. While the features are mutually interdependent in the search process, exact importance of each feature and its dependencies should be subject of future research such as in the context of information retrieval end-user studies.

Table 1 below lists the features further mapped to the LRM user tasks of finding, identifying, selecting, obtaining and exploring. The plus sign (+) applies when the feature is

		Find	Identify	Select	Obtain	Explore
1	Browsing by subject access points: subjects from controlled vocabularies, e.g. subject headings, captions from classifications systems, free keywords	-	-	-	-	+
2	Searching by subject access points from controlled vocabularies, including by individual words	+	+	-	-	-
3	Browsing by facets, aspects and individual concepts from controlled vocabularies, such as individual terms from subject headings, as well as captions and notations representing individual concepts from synthesized classmarks (e.g. in Universal Decimal Classification)	-	+	+	-	+
4	Searching by any combination of individual concepts and facets (as above)	+	+	+	-	-
5	Searching by major and minor themes represented by controlled vocabularies, if supported by the indexing policy	+	+	-	-	+
6	Presenting and browsing excerpts of concept hierarchies (e.g. a classification scheme, a thesaurus), matching words and phrases from search terms, including for disambiguation, narrow, broader and related searching	+	+	-	-	+
7	Auto-completing search terms once the user begins typing	+	+	-	-	-
8	Auto-suggesting authorized controlled versions of entered search terms, presenting all the relationships and allowing further choice in browsing or searching the controlled vocabularies	+	+	-	-	+
9	Suggesting corrections of mistypes	+	+	-	-	-
10	Searching by words from various metadata elements and full text	+	-	-	-	-
11	Combining controlled subject searching with searching by other bibliographic fields	+	+	-	-	-
12	Highlighting search terms in retrieved metadata and resources	+	+	-	+	-
13	Advanced searching by Boolean and proximity operators, truncation of searches, wildcard searches	+	-	-	-	-
14	Linking each subject access point to its resources	-	-	+	+	+
15	Linking subject access points from one controlled vocabulary to corresponding concepts in others	-	-	-	-	+
16	Adding, browsing and searching end user tags	+	+	-	-	+
17	Combining previous search formulations	+	-	-	-	-
18	Help on searching	-	-	-	-	-
19	Searching by image-related features [e.g. adding search criteria to retrieve only results/records with or without corresponding image representation; image orientation (portrait or landscape); image size]	+	+	+	+	-
20	Searching by <i>Content-based Image Retrieval (CBIR)</i> methods (e.g. query by example image; query by sketch map; query by colour map)	+	+	+	+	-
21	Features related to implementation of the International Image Interoperability Framework (IIIF) open standards for enhanced user experience (e.g. deep zoom viewing, comparing, manipulating and annotating images, API interfaces)	-	+	+	+	-

Table 1.
Mapping subject access features to LRM tasks

relevant to the related task and the minus sign (–) applies when the feature is not relevant to the related task. This is our initial proposal calling for further research. What we should be especially aware of are differences ensuing from situations when end-users’ activities are performed against authority data and those when activities are performed against metadata related to resources. For example, LRM distinguishes the following two situations related to the “obtain” task: (1) to obtain a resource by linking to or downloading an online resource using the link found in the library catalogue; and, (2) to obtain information about an entity itself from the information recorded in authority data.

3. Methodology

3.1 Purpose and aims

With the purpose of helping inform future policies on access to cultural heritage in Sweden and beyond, the study aims to determine the current status of online search services providing access to cultural heritage of Swedish museums. Specifically, this study investigates to what degree these online search services support end-user subject searching against a set of 21 criteria (see the section “Desirable features of online subject access” above).

3.2 Sample and method

All of Sweden’s museum websites and cross-search services known to the authors were examined (archives and libraries were left out due to limited resources available for the study but should be covered in future research). The museums were identified in a process comprising several steps. We first used Wikipedia’s list of museums as a starting point (Wikipedia, 2020). Through this step, we identified 132 museums, of which 45 provided online access to at least a part of their collections. A number of these museums linked to additional museums not identified via the first step, which were then added to the list. This updated list was in the third step complemented by Google Web search, resulting in the total of 107 additional museums not identified in the first step. The very final list comprised 239 museums, of which most were identified as having no online access to their collections and provided only a web page with basic information such as visiting hours and location; 60 of those were actually heritage sites and could be considered in their entirety as information objects (e.g. Ransåters bruksherrgård, a family estate of Erik Gustaf Geijer; or, Motala longwave transmitter, a broadcasting station). A few museums’ web pages were not accessible at the time of research, so they were also excluded from the sample. The final number of museums identified as having online access to at least some part of their collections was 91, which represents the research sample. In addition, when exploring the online museums’ websites, some linked to their own collections in external, cross-search services. A total of nine different cross-search services were identified and also added to the sample which in summary comprised 91 individual museums’ websites and nine cross-search services.

The total of these 100 search interfaces were examined in the period from 1 July to 31 September 2020. Each online service was examined by one of the authors of the paper and the data observed were recorded in a spreadsheet. The data recorded were name of the institution; type of service (cross-search or individual); search interface URL; absence or presence of each of the 21 features (see section “Desirable features of online subject access”) with notes describing implementation if present; and other notes on the search tool(s).

4. Results

4.1 A general overview

Out of the 91 museums in Sweden providing online search services to their collection, the majority (74 or 81.3%) make their collections searchable via cross-search services; of those, 15

museums (16.5%) use cross-search services as well as their own search tool, while the remaining 17 (18.7%) use only their own search tool only. The cross-search services used, described further below, are the following: Kringla (62 museums), Europeana (60), Digitalt Museum (38), Carlotta (14), Wikimedia Commons (4), eMuseumPlus (3), Alvin (2), Google Art Project (2), and Musical Instrument Museums Online (MIMO) (1). Of museums which use their own search tools (either on their own or in combination with a cross-search interface), seven museums have chosen the same interface, known as Kulturhotell. In addition, Stockholm City Museum and the Museum of Medieval Stockholm both use the same interface; the two museums of modern art do likewise: Moderna Museet in Stockholm and Moderna Museet in Malmö both use the same interface.

The most commonly used cross-search services are Kringla, Europeana, Digital Museum and Carlotta. Kringla (<http://www.kringla.nu/kringla/>) is a Swedish cross-search service managed by the Swedish National Heritage Board (Riksantikvarieämbetet). Kringla is an end-user interface to the Swedish Open Cultural Heritage (SOCH) web service and aggregator, which collects data from 74 Swedish institutions, of which 62 museums are in the current study sample, and which contributes to Europeana (Riksantikvarieämbetet, 2019). Europeana (<https://www.europeana.eu/en>) is an international, digital search interface of European cultural heritage providing access to thousands of archives, libraries and museums across the continent. The Digitalt Museum (<https://digitaltmuseum.org/>) interface is financed by Arts Council Norway with the purpose of making Norwegian and Swedish museums' collections freely available on the Internet (DigitaltMuseum, 2021) although this seems to be contingent on buying the Primus collections management system. Carlotta (<http://carl.kulturen.com/web>) is an information system for museum collections managed by the Swedish State Museums of world culture (Museum of World Culture; Ethnographic Museum; Museum of Far Eastern Antiquities; The Museum of Mediterranean and Near Eastern Antiquities). It is designed with flexibility in mind and is based on CIDOC CRM but adapted to the Swedish context (Kulturen, 2021).

Less commonly used cross-search services are Wikimedia Commons, eMuseumPlus, Alvin, Google Arts and Culture, and MIMO. The eMuseumPlus (<http://emuseumplus.lsh.se/eMuseumPlus>) platform is owned by three Swedish museums of history or, in Swedish, Statens historiska museer (SHM): Livrustkammaren, Skoklosters Slott and Hallwylska Museet. For these, the eMuseumPlus works as a cross-search service; however there are three more museums using the same search engine (National Gallery, Gothenburg Museum of Art and Röhsska Museum), which are not cross-searchable via the platform (eMuseumPlus, 2021a, b). Wikimedia Commons (<https://commons.wikimedia.org>) is an international and multilingual media file repository created and maintained by volunteers, and based on wiki-technology (Wikimedia Commons, 2021). Alvin (<http://www.alvin-portal.org>) is a Swedish platform and catalogue intended to be used for the long-term preservation and accessibility of digitized material from Swedish cultural heritage institutions. It is developed and maintained by Uppsala University Library and run as a consortium with Gothenburg University Library and the University Library at Lund University (Alvin, 2021). Google Arts and Culture (<https://artsandculture.google.com/>) is Google's non-profit initiative that aims to preserve and give access to digital or digitized cultural objects from all over the world (Google, 2021). MIMO (<https://mimo-international.com/MIMO/>) is a freely accessible database of information about musical instruments held in public collections. MIMO was started by five European cultural and educational institutions and was established by the European Commission. MIMO associates non-specialist vocabulary with terms and classification systems used by professionals (MIMO, 2021).

Judging already from the descriptions of the cross-search services found on their respective websites, we note that only two of them, Carlotta and MIMO, mention any museum standards discussed in the Background section.

The 21 subject search features (see the section “Desirable features of online subject access”) are in general rarely used. This holds true both for online search services with their own search tools and for cross-search services; these results are similar to those reported for bibliographic databases (Golub *et al.*, 2020) and for discovery services (Golub, 2018).

Of the 29 museums with their own search tools, 1 museum used as many as 10 search features (Sörmland Museum), 5 museums used 8 features (5 out of 7 using Kulturhotell), 2 museums used 7 features (2 others using Kulturhotell), 3 museums used 6 features (3 individual eMuseumPlus museums: the National Gallery, the Gothenburg Museum of Art and Röhsska Museum), 1 museum used 5 features (the Swedish Museum of Performing Arts), 1 museum used 4 features (the Swedish Museum of Natural History), 4 museums used 3 features (the Swedish History Museum, Naturhistoriska museum and both Moderna Museet museums), 6 museums used 2 features (Gustavianum, the Paleontological Museum of Uppsala University, the Museum of Work, Stockholm County Museum and both Digitala Stadsmuseet museums), 3 museums used only 1 feature (Industrimuseum, Gotland Museum, Litografiska Museet) and 5 other museums did not use any feature [Västernorrlands Museum, the Museum of Sketches for Public Art, Arlanda flygsamlingar (Arlanda Civil Aviation Collection), Waldemarsudde, Teleseum, the Thiel Gallery]. Of the latter 5 institutions, 3 (Västernorrlands Museum, the Museum of Sketches for Public Art, Waldemarsudde) are not present in any cross-search services either, meaning that online access to their collections hardly exists due to the lack of any subject search support. The features which are used most, described in the order of frequency by feature number, are 10 (17 museums), 18 (11 museums), 13 (12 museums), 2, 7 and 14 (11 museums), 1 (10 museums), 11 (7 museums), 17 (3 museums), 12 and 19 (2 museums). Features 3, 4, 5, 6, 8, 9, 15, 16, 20 and 21 were not used at all.

The use of features in cross-search services is identified as follows: 1 used 7 features (Wikimedia Commons), 3 used 6 features (Carlotta, eMuseumPlus and Europeana), 3 used 5 features (Alvin, Digitalt Museum and Kringla), 1 used 4 features (Google Art and Culture) and 1 used a single feature (MIMO). The features which are used most, described in the order of frequency by feature number, are 7, 13 and 18 (6 cross-search services), 1 and 10 (5 cross-search services), 17 and 19 (4 cross-search services), 9, 11, 12, 20, 21 (2 cross-search services) and 2 and 14 (1 cross-search service). Features 3, 4, 5, 6, 8, 15 and 16 were not used at all.

Table 2 below presents an overview of the functionalities across all of the 100 websites examined; each functionality is then discussed in more detail in the following section. Please note that Table 2 shows the number of museums rather than cross-search services; thus, the number of cross-search services using a feature is multiplied by a number of museums using each cross-search service. Some cross-search services are used by many museums in the sample, others are used by only a few of them. For example, the first search feature, a very common one (browsing by subject access points: subjects from controlled vocabularies), is used by 10 museums’ own search tools (Sörmland Museum, Museum of Work, Motala Industry Museum and all 7 museums using Kulturhotell’s interface) and 5 cross-search services, which are Digital Museum (representing 38 museums), Europeana (60 museums), Wikimedia Commons (4 museums), Google Arts and Culture (2 museums) and MIMO (1 museum). One should also be aware of the fact that many museums are present on more than just one cross-search service, that is, Livrustkammaren is present on Europeana, Wikimedia Commons and Google Arts and Culture, but in the study, the museum is counted only once. Livrustkammaren is present even on other cross-search services, and it uses its own search service but those do not support the mentioned search feature. In all, this is why the number of museums present on the 5 cross-search services using the first search feature is only 63 and the total number of museums using the feature is 73.

	Museums' own search tools using it	Cross-search services using it	Number of museums' collections using it	Ratio	
1	Browsing by subject access points: subjects from controlled vocabularies, e.g. subject headings, captions from classifications systems, free keywords	10	5	73	80.2%
2	Searching by subject access points from controlled vocabularies, including by individual words	11	1	13	14.3%
3	Browsing by facets, aspects and individual concepts from controlled vocabularies, such as individual terms from subject headings, as well as captions and notations representing individual concepts from synthesized classmarks (e.g. in Universal Decimal Classification)	0	0		
4	Searching by any combination of individual concepts and facets (as above)	0	0		
5	Searching by major and minor themes represented by controlled vocabularies, if supported by the indexing policy	0	0		
6	Presenting and browsing excerpts of concept hierarchies (e.g. a classification scheme, a thesaurus), matching words and phrases from search terms, including for disambiguation, narrow, broader and related searching	0	0		
7	Auto-completing search terms once the user begins typing	14	6	75	82.4%
8	Auto-suggesting of authorized controlled versions of search terms, presenting all the relationships and allowing further choice in browsing or searching the controlled vocabularies	0	0		
9	Suggesting corrections of mistypes	0	2	66	72.5%
10	Searching by words from various metadata elements and full-text	20	5	83	91.2%
11	Combining controlled subject searching with searching by other bibliographic fields	7	1	9	9.9%
12	Highlighting search terms in retrieved metadata and resources	2	2	65	71.4%
13	Advanced searching by Boolean and proximity operators, truncation of searches, wildcard searches	15	6	81	89%
14	Linking each subject access point to its resources	11	1	25	27.5%

(continued)

Table 2.
A summary of functionalities across all the websites

		Museums' own search tools using it	Cross-search services using it	Number of museums' collections using it	Ratio
15	Linking subject access points from one controlled vocabulary to corresponding concepts in others	0	0		
16	Adding, browsing and searching end user tags	0	0		
17	Combining previous search formulations	6	4	63	69.2%
18	Help on searching	16	6	80	87.9%
19	Searching by image related features (e.g. adding search criteria to retrieve only results/records with or without corresponding image representation; image orientation (portrait or landscape); image size)	6	4	72	79.1%
20	Searching by <i>Content-based Image Retrieval (CBIR)</i> methods (e.g. query by example image; query by sketch map; query by colour map)	0	2	60	65.9%
21	Features related to implementation of the International Image Interoperability Framework (IIIF) open standards for enhanced user experience (e.g. deep zoom viewing, comparing, manipulating and annotating images, API interfaces)	0	2	4	4.4%

Table 2.

4.2 Analysis against the 21 desirable features

This section presents results for each of the 21 functionalities across all individual and cross-search online services. Each functionality is listed by name with an analysis, using an example with a screenshot to illustrate interesting features.

4.2.1 Feature 1: *Browsing by subject access points: subjects from controlled vocabularies, like subject headings, captions from classifications systems, free keywords.* This feature is used by 10 museums' in-house search tools and 5 cross-search services, covering 73 museums in the sample (80.2%). Of the in-house search tools, Sörmland Museum's tool for browsing (<https://www.sorlandsmuseum.se/utforska/>), under Subject, allows the user to choose between 17 different broad categories by which all the objects in the museum collection are classified (e.g. love, war, clothes, etc.), rather than supporting subject browsing by concepts from a controlled vocabulary (<https://www.sorlandsmuseum.se/utforska/?TypeCategory=&SubjectCategory=51&SortCategory=index&SearchText=>). Similarly, the Museum of Work allows browsing using several simple categories related to motifs such as Swedish politicians, communism, equality, etc. (<https://www.arbetetsmuseum.se/ewk-museet/sok-i-ewk-databas/>). Likewise, in the Motala Industry Museum, the only way to explore the museum's collection is to browse through over 12,000 of its photographs. The user can choose one of 28 broad categories which are not from a controlled vocabulary but list topics like bridges, boats, employees, history, etc.; no search option is available (http://www.motala-industrimuseum.com/index_bild.php).

Figure 2 below shows what browsing features look like:

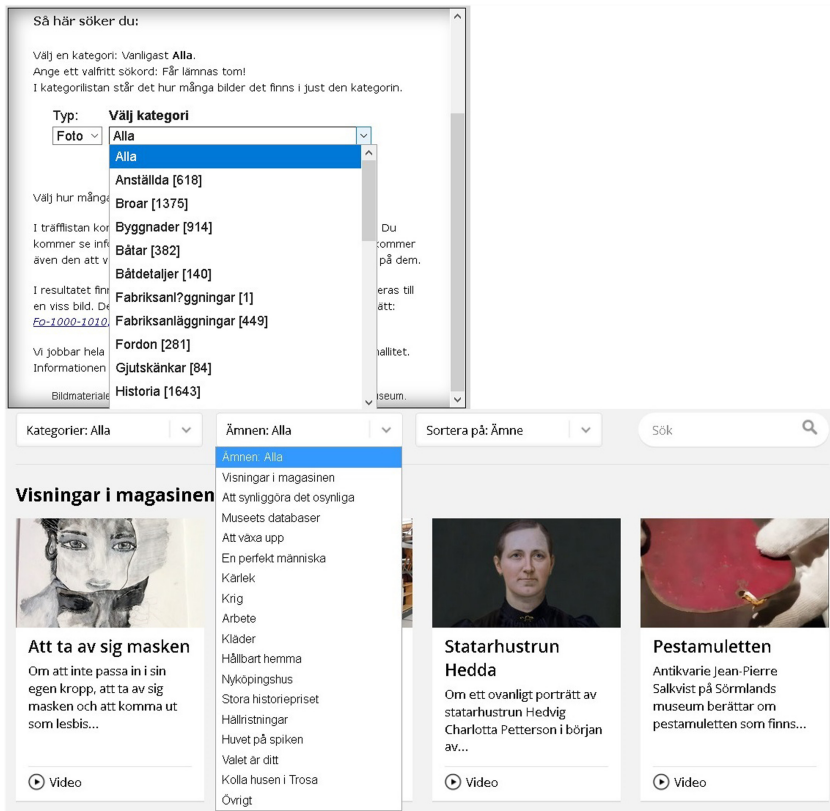


Figure 2. Subject browsing in individual museums using a small number of subject categories

As in Figure 3 below, museums using Kulturhotell are different because of a much larger set of categories – or rather keywords used to describe museum objects (“Bläddra efter nyckelord”); there are over 17,000 keywords, and browsing is allowed based on an



Figure 3. Alphabetical subject browsing of keywords

alphabetical order, number of objects and date when the object was added (https://blm.kulturhotell.se/items/tags?per_page=72). However, an alphabetical listing of so many keywords is not really user friendly; hierarchical subject browsing is much more suitable.

Of cross-search services, in Digitalt Museum, the user can seemingly browse through more than 220,000 topics although actually only 100 of the most common topics (those with most documents) are accessible for viewing and browsing (<https://digitaltmuseum.org/search/?aq=>): see Figure 4 below. As in the previous example, only 100 keywords, ordered only by popularity will not suffice when there are over 220,525 topics available for over 6,000,000 objects.

Europeana allows alphabetical browsing of topics (<https://www.europeana.eu/en/collections/topics>) resulting in 19 pages of categories (24 categories per page). No controlled vocabulary seems to be in use.

Google Arts and Culture, among many browsing options (e.g. by colour, popular topics, occasional topics, e.g. Easter or Christmas), supports browsing by six broad categories: artists, mediums, art movements, historic events, historical figures, and places, of which the last four are subject related. The categories are further subdivided into more specific topics (artists into artist names); at the second hierarchical level, very large categories can be listed alphabetically or in a timeline; specific objects by an artist can be ordered by popularity, time and colour (<https://artsandculture.google.com/explore>).

Wikimedia Commons supports browsing by an elaborate, multi-level hierarchical classification of topics with many levels (https://commons.wikimedia.org/wiki/Main_Page). The classification based on topics is structured into four main categories at the first level: (1) nature; (2) society culture [sic]; (3) science; and (4) the environment; these are further subdivided into a number of more specific hierarchical levels. This is closest to what a hierarchical browsing structure should look like, as the user learns about the concept space and acquires knowledge about the world based on how the classification presents it. Such a classification tree should ideally also have a browsable number of items – a few dozens, not hundreds or more as we often see in online search services.

Overall, it is important to observe that the lists of topics in the small number of services providing this feature do not seem to be taken from any kind of controlled vocabularies but are adjusted to suit the museum collections. The reason for this could be the fact that there are no

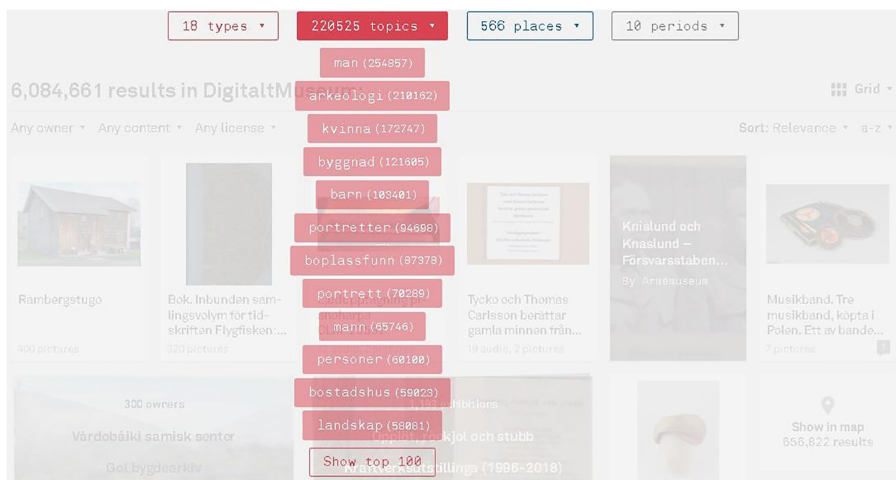


Figure 4. Browsing based on a hundred of the most popular keywords in a cross-search service

national authority files available that would allow for subject vocabulary control and/or the fact that dominant English vocabularies such as AAT mentioned above have not been translated into Swedish. However, this situation prevents interoperability across collections and also puts unnecessary demand on the user to learn about new structures for every museum.

Furthermore, when providing options for subject browsing, the rationale is to provide the end users with an insight into the subject area; subject browsing is particularly useful when the user does not know what specifically to look for or is new to the subject area or museum and is therefore unable to form a good search query. The cross-search service eMuseumPlus states in its instructions that a classification system that it uses for object types (although not for subjects) could be a good tool to gain an overview of the collections or assist when one does not really know what one is looking for (eMuseumPlus, 2021a, b); this demonstrates that information scientists are aware of this. However, having only a small number of subject categories for a large number of objects is not very meaningful for the user; having only one or two dozen categories for hundreds or even thousands of objects will again result in a list which is of little use as few users would take the time to go through a long list of objects from the same category. Between a dozen and a few dozen objects per subject would be more practical for the user, and this is why multi-level detailed hierarchical classification systems such as those used by Wikimedia Commons are useful. If, in addition, they could be standardized and used across different museums, that would be ideal so that the user would need to learn about the specific concept space only once. The classification systems need to be developed on the basis of identified user requirements and updated regularly, at the same time following international standards and guidelines for creating controlled vocabularies (see also the section Background above).

4.2.2 Feature 2: Searching by subject access points from controlled vocabularies, including by individual words. The feature is used by 11 museums' in-house search tools (the Swedish Museum of Performing Arts, the Swedish History Museum, Stockholm County Museum, Sörmland Museum and 7 Kulturhotell) and 1 cross-search service (Alvin): 13 museums in total (14.3%). Of the former, the Swedish Museum of Performing Arts, which uses Axiell software, allows searching by many metadata elements. One that is subject related is a motif of person (<https://calmview.musikverk.se/CalmView/Advanced.aspx>), which gives the appearance of using an authority file; however, it is hard to say if it really uses one (no information is given there or under Help). The Swedish History Museum allows the user to search by 10 "topic-like" object categories (e.g. religion and cult, transport) (<http://mis.historiska.se/mis/sok/sok.asp?qtype=f&page=4>), as in Figure 5 below. Similarly, Stockholm County Museum allows the user to refine a search using six general categories (everyday, typical of the time, communication, lifestyle, change and my place) (<https://samlingar.stockholmslansmuseum.se/>).

Users of *Sörmland Museum*, here using its advanced search tool (among many other search options) can search by a motif related to a person or a place shown in a picture (Figure 6 below).

Figure 7 shows searching by person in the photograph. The field for a person's name seems to be taking data values from an authority file (<https://sokisamlingar.sormlandsmuseum.se/items/search>).

As in Figure 8 below, all the museums that use Kulturhotell (<https://blm.kulturhotell.se/items/search>) allow the user to search by several subject-related metadata fields: 5 different subject fields (ämne) and 14 motif-related categories (mostly for location but even motif category); 5 of 7 museums using Kulturhotell even support searching by person in the photograph ("personer i bild" in Swedish).

Of the cross-search services, Alvin supports searching by subject; the field is reportedly controlled (SAO – Swedish Subject Headings and TGMII – Thesaurus for Graphic Materials II) but no list of terms is shown in the search tool, so the user cannot know which terms can be used.

- föremål

Skicka förfrågan | Hjälp | Translär

Sök: Inventarienummer | **fyndplatser**

Sök bland Historiska museets **föremål**. Vi h uppdateras fortlöpande. Ange ett eller flera k

Vad

Samlingsområde:

Föremålskategori: **Välj kategori**

- Vapen, rustningar och fångstredskap
- Hantverk och redskap
- Husgeråd och livsmedel
- Religion och kult
- Byggnadsdetaljer och monument
- Klädedräkt och personlig utrustning
- Nöjen och laster
- Handel och värdemätare
- Transport och samfärdsel
- Vetenskap och rättsväsende

| Kategorier»

Sakord: Lista sakord

Typ: Lista typer

Del: Lista delar

Material: Lista material

Inventarienummer:

Undernummer:

Fyndnummer:

temåsbilder | Taggar

tabas (ca 21 % av det beräknade antalet) och databasen

Figure 5. Searching by 10 subject categories in the Swedish History Museum

Sök på samling

*Tips! Håll in **Ctrl** eller **⌘**-knappen för att välja flera fält samtidigt.*

Sök på typ

*Tips! Håll in **Ctrl** eller **⌘**-knappen för att välja flera fält samtidigt.*

Personer i bild

- Blomsjö, Anna** på fullständigt namn.
Titel: Fru
- Blomsjö, Anna Lisa** på fullständigt namn.
Titel: Fröken
- Celsing, Anna-Stina, von** på fullständigt namn.
Titel: Fröken
Ort: Bilby
- Enlöv, Anna-Stina** på fullständigt namn.

Utökad kriterier

Välj nedan

Kriterium kan förändras beroende på om du söker på namn på person i bild

Relationstyp

Välj nedan

Har inte relation:

Figure 6. The Sörmland Museum user interface providing a controlled search term for people as subjects

Personer i bild

- Andreas** på fullständigt namn.
- Anna** på fullständigt namn.
Ort: Hörvik
- Adolfs, Hanna** på fullständigt namn.
- Anderson, Hugo** på fullständigt namn.
Titel: Länsjägmästare
- Andersson, Elsa** på fullständigt namn.

Utökad kriterier

Välj nedan

Kriterium kan förändras beroende på om du söker på namn på person i bild

Relationstyp

Välj nedan

Har inte relation:

Figure 7. Kulturhotell, searching by person in the photograph

Objekttyp	Fotografi
Samling	• Sydbilds arkiv
Nyckelord	Handelshamn, Militär, <u>Uniform</u>
Bildnr / samlingsnr / inventariernr	Sba 19691029 d 03
Ämne 1	Båt
Ämne 2	Hamn
Ämne 3	Minfartyg
Ämne 4	Marinen
Ämne 5	Flotta
Tillkomsttid start	1969-10-29
Motivkategori	Porträtt
Motiv - Land	Sverige
Motiv - Landskap	Blekinge
Motiv - Län	Blekinge
Motiv - Kommun	Karlskrona
Motiv - Härad	Östra

Figure 8.
Subject searching in Kulturhotell

Similarly, while Carlotta supports searching by motif (“motivkategori” in the Swedish interface), it is not clear where the terms are coming from and whether they are controlled (<http://carl.kulturen.com/web>). DigitaltMuseum allows for searching by topic in its advanced search tool, but as in the former case, it is unclear if the terms are controlled (<https://digitaltmuseum.org/search/advanced>). Therefore these two search services are not counted amongst those implementing this feature.

Of the museums with their own search tool, the Swedish Museum of Performing Arts uses Axiell’s Calmview Advanced Search interface which allows searching by a field called “keyword” obviously not from a controlled vocabulary, so this cannot be counted (<https://calmview.musikverk.se/CalmView/advanced.aspx?src=CalmView.Catalogue>). Of the cross-search services, Alvin supports searching by subject in its Extended Search interface, but it does not specify whether the values are controlled and how to search (<http://www.alvinportal.org/alvin/advanced.jsf?dswid=-6194&searchType=EXTENDED&query=&aq=%5B%5B%5D%5D&aqe=%5B%5D>). The Carlotta software (e.g. Kulturen at <http://carl.kulturen.com/web>), under Simple Search, provides a search by “name” but does not further specify whether the terms are taken from a controlled vocabulary, which implies that they are not. Advanced search is only for museum specialists since it supports searching by database fields using specialized terms such as FOLNAM, OBJIDN, etc. The Help page does not specify any use of KOS. The interface of eMuseumPlus (<http://emuseumplus.lsh.se/eMuseumPlus?service=ExternalInterface&module=collection&moduleFunction=search>) supports searching by person (field: “namn”) but does not use any controlled vocabulary. Kringla supports subject searching in its Detailed Search in two user-friendly named fields “What are you looking for?” and “Object title”; however, these do not seem to be taken from a KOS (<http://www.kringla.nu/kringla/>). So this cannot be counted either. Moderna museet allows subject searching in its Advanced Search through its Title field (<https://sis.modernamuseet.se/sv/advancedsearch>), but there are no KOS-based fields either.

4.2.3 *Features 3 to 6.* Features 3, 4, 5 and 6 take further advantage of characteristics of controlled vocabularies to support the end user in subject searching. Feature 3 is Browsing by facets, aspects and individual concepts from controlled vocabularies, such as individual terms from subject headings, as well as captions and notations representing individual concepts from synthesized classmarks (e.g. in Universal Decimal Classification). This would allow browsing by specific topics from existing categories and would require the use of advanced browsing interfaces: none of the interfaces support this. Feature 4, searching by

any combination of individual concepts and facets (as outlined in the preceding features) would allow very specific query formulation and retrieval of highly precise results. This feature requires, however, that suitable, quality-controlled controlled vocabularies such as AAT (see Background above) are actually used. Wikimedia Commons allows searching by words using its categories. Feature 5, searching by major and minor themes represented by controlled vocabularies, if supported by the indexing policy, does not seem to be supported by any of the search tools. If it was, it would allow high precision in the retrieval because it would be possible for the user to specify whether a subject index term is of major or minor importance for the document at hand. Feature 6 includes presenting and browsing excerpts of concept hierarchies (e.g. a classification scheme or a thesaurus), matching words and phrases from search terms, including for disambiguation, narrower, broader and related searching. It again builds on subject indexing based on high-quality controlled vocabularies. None of the search tools in the study use this feature; if they did, the end user would be able to disambiguate a search term, for example, whether the term “bank” is related to banks as financial institutions or to banks of a river; and at the same time it would be able to look for narrower terms or broader terms as presented within a hierarchy of related concepts. Although Wikimedia Commons has an elaborate classification tree of categories, it does not support searching by categories except where relevant excerpts from the classification tree are retrieved for further search query formulation.

4.2.4 Features 7, 8 and 9. The following three features are closely related: feature 7, auto-completing search terms once the user begins typing; feature 8, auto-suggesting authorized controlled versions of entered search terms, presenting all the relationships and allowing further choice in browsing or searching the controlled vocabularies; and feature 9, suggesting corrections to mistypes. In the study sample, since few controlled vocabularies are used, none of the search tools seem to provide automatic translation into controlled terms (feature 8).

The auto-suggest feature proposes the user a list of suggested terms after the user has started typing a search phrase. The feature is used by 14 museums’ in-house search tools (the Swedish Museum of Natural History, Sörmland Museum, both Moderna Museet museums, all 7 Kulturhotell and 3 individual eMuseumPlus museums), and 6 cross-search services (eMuseumPlus, Europeana, Kringla, Google Art and Culture, Wikimedia Commons, MIMO), a total of 75 museums (82.4%).

This feature works in a similar way in all the search services: the user is given a list of matching terms immediately after having typed the first few letters of a search term in the search box (as shown above in [Figure 9](#), which shows its use in Alvin); however, 1 out of 2

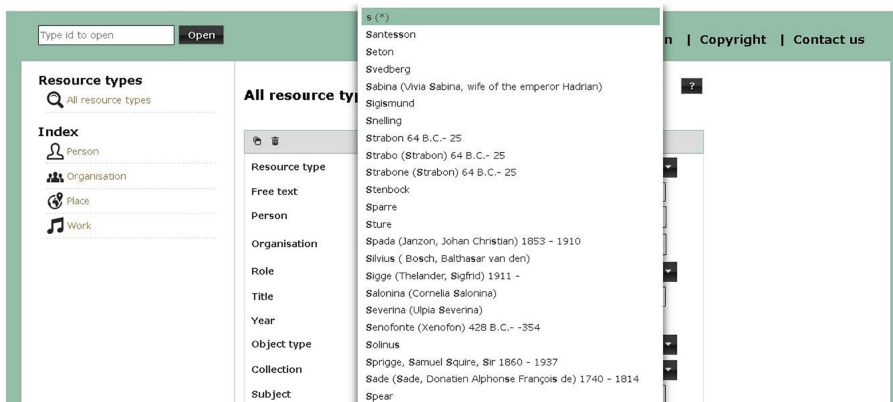


Figure 9.
Autocompletion of
search terms in Alvin

Sörmland Museum’s search tools retrieves records directly where the search string begins the words in the records retrieved. [Figure 10](#) from Sörmland Museum shows how, upon typing the string “ann”, records will be retrieved which include this string in the names of objects or other metadata.

Sörmland Museum, Kulturhotell (5 Kulturhotell museums of 7 use this feature), Alvin and eMuseumPlus are the only services using controlled vocabularies in this way, while all the others use free keywords.

None of the museums or cross-search services provide feature 8, that is, the auto-suggesting function also presents related term relationships and allows further choice in browsing or searching the controlled vocabularies. However, the Swedish Museum of Natural History seems to retrieve matches which are not simply string matches although it is not clear how these suggestions are derived. As in [Figure 11](#) below, if one starts typing the string “hun” under Classification, a drop-down list of terms containing the string appears, showing it in bold (e.g. “Hunterius swedenborgii” for a right whale, “Empria hungarica” for a parasite, etc.)



Figure 10. Autocompletion of both controlled and non-controlled terms in Sörmland Museum

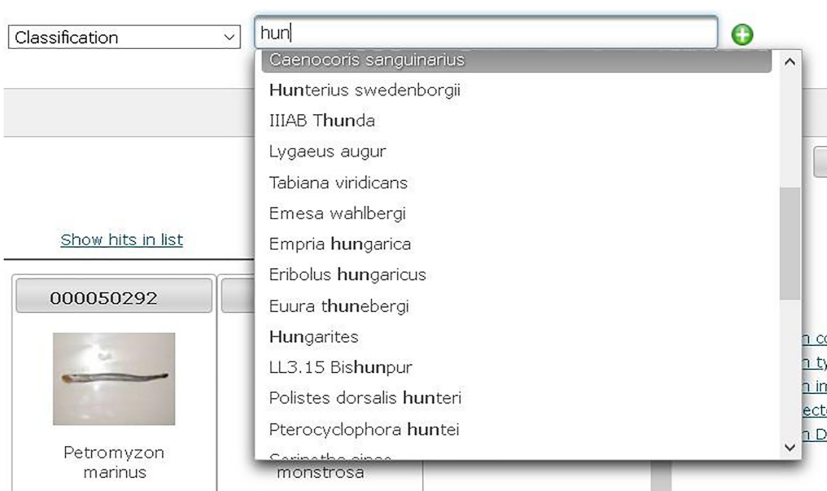


Figure 11. Auto-suggestion results in the Natural History Museum

while also listing terms which do not match the search string but are synonyms (e.g. the insect “*Caenocoris sanguinarius*” is a synonym of “*Thunbergia sanguinarius*”).

Of the cross-search services, MIMO is a good example of providing multilingual auto-suggestion in their simple search box. While Kringla also supports auto-suggestions, it also seems to suggest misspelled versions. Figure 12 below shows how, when entering “annn”, it will suggest misspelled versions with three “n” letters instead of two such as “annnat” (likely a misspelling for “annat” meaning “other”), “annnan” (likely a misspelling for “annan” meaning also “other”), and “annnars” (likely a misspelling for “annars” meaning also “otherwise”).

Feature 9 supports suggesting corrections for mistypes. None of the museums’ own search tools seem to support this; two cross-search services do (66 museums, i.e. 72.5%). Kringla suggests the right word and asks “did you mean . . .?”, but as in Figure 12 above, there are some objects that match the mistyped search string, in which case the suggestions are not corrected. Wikimedia Commons has implemented the feature in the same way, with “did you mean . . .” as in Figure 13 below.

Figure 12.
An example of how autosuggestions in Kringla occasionally provide erroneous suggestions

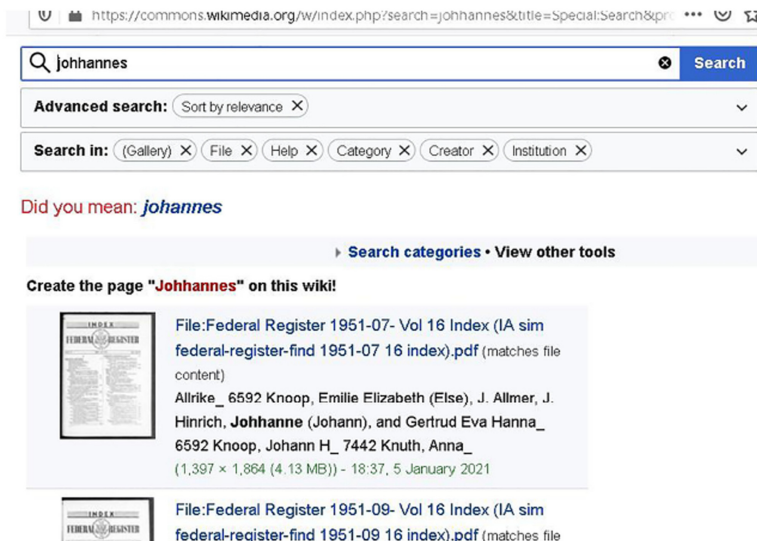
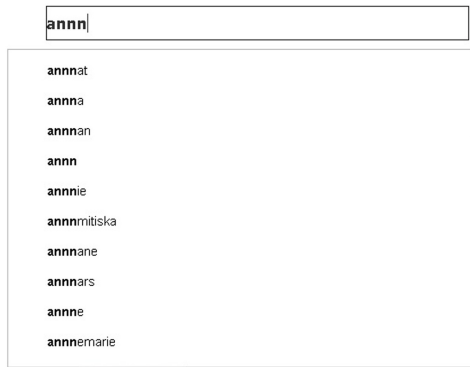


Figure 13.
Wikimedia Commons example of suggesting corrections for mistypes

4.2.5 *Features 10, 11, 12, 13, 14.* The following features are all related to searching. Feature 10, searching by words from various metadata elements and full-text is rarely implemented in museums because museum objects are rarely full-text documents. Instead, in this context, we take full-text to mean full metadata records, as searching on all metadata fields will improve recall. There are 20 museums with their own search tools and 5 cross-search services (83 museums, i.e. 91.2%) that allow searching through many different metadata fields. These are Kulturhotell (>100 metadata fields), Sörmland Museum (33 fields), Alvin (17), Carlotta (15), the Swedish Museum of Natural History (13), the Swedish History Museum (12), DigitaltMuseum (11), the Paleontological Museum of Uppsala University (10), the Swedish Museum of Performing Arts (9), eMuseumPlus (7), Kringla (7), Moderna Museet (6), Naturhistoriska museum (5), Gotland Museum (4) and Litografiska Museet (4). The highest number of possible metadata fields combined in one search is returned by Kulturhotell, which allows searching by over a hundred [sic] metadata fields in one search; however, it is doubtful if any of the objects in their collections is described using all of the fields.

Feature 11, combining controlled subject searching with searching by other bibliographic fields, is supported by those which are listed under feature 10, but which at the same time use at least one controlled field. There are 7 museums whose in-house search tools are in this group (the Swedish Museum of Performing Arts, Sörmland Museum and 5 out of 7 museums using Kulturhotell) and 1 cross-search service (Alvin), giving a total of 9 museums (9.9%). The controlled subject fields have been already described under feature 2.

Feature 12, highlighting search terms in retrieved metadata and resources, is implemented by two museums in their own search tools and two cross-search services (i.e. 65 museums or 71.4%). The Swedish Museum of Performing Arts, using Axiell software, highlights search terms in the retrieved metadata and resources (both in the simple and advanced search); however, the search term is highlighted only when the metadata record is open rather than in the list of retrieved snippets, as in [Figure 14](#) below.

Sörmland Museum highlights the search string in the retrieved list of snippets but not in the open record. Both options are required for the best user experience. The museum allows this option only in the simple search interface, but it would be better to offer it in both the simple and advanced interfaces.

Of the two cross-search services, Wikimedia Commons highlights search terms in the results list and Kringla has the best implementation, both in the results list and after opening a selected metadata record.

Feature 13, advanced searching by Boolean and proximity operators, truncation, and wildcard search, is used in different ways by 15 museums in their own search tools and 6 cross-search services (i.e. 81 museums or 89%). A textbook example of the use of Boolean operators can be found in the Swedish Museum of Natural History, which allows combining the input of several search strings via different search boxes, which are then combined with Boolean operators (AND, OR, NOT) as in [Figure 15](#) below.

In other museums the user is expected to write the Boolean operators in the search box. Interfaces which support this kind of implementation are

- (1) Digitala Stadsmuseet (<http://digitalastadsmuseet.stockholm.se/fotoweb/>), which requires the user to type “AND” (to retrieve documents matching both searched terms), “OR” (to retrieve document matching any of the search terms), “NOT” (to exclude a term from searching); it also supports the use of “*” for truncation (to search for all forms of a word) and “?” for a wildcard (a symbol that can represent any character).
- (2) Kulturhotell and Sörmland Museum, which requires the use of “+” for AND, “-” for NOT, “*” for truncation, and quotation marks for searching documents that match the exact phrase.

About the catalogue	Advanced Search > Search Results > Record	
▼ Help	First Previous 6 of 142	
How do I search the database?	<input type="checkbox"/> STM 2007-1062 a	
What are acceptable searches?	Förvarande institution	Scenkonstmuseet
What is in the catalogue?	Refkod	STM 2007-1062 a
Glossary	Inventarienummer	STM 2007-1062 a
Send us feedback	Sakord/Objektnamn	Plakett
	Beskrivning	plakett för "Årets Svensk Amerikan" rektangulär, Vasa Orden av Amerika vapen: blå emaljerad malteskors och guldfärgad vasa och guldbokstävar "V"och "O" i metall fastsatt på övre halvan; på nedre halvan gravyr "ÅRETS/ SVENSK-AMERIKAN/ 1989/ SIGNE HASSO/ DISTRIKTLOGERNA/ NORRA SVERIGE NR 19/ SÖDRA SVERIGE NR 20". Hela framsida omramad av en bladrelief. Rött och blå passepartout klistrad på baksida, samt guldrad över hörnorna som är klistrad fast vid baksida
	Anmärknningar	utan passepartout , Hasso, Signe
	Teknik	relief
	Material	mässing

Figure 14. Highlighting of a search term in an open retrieved record only

Naturarv

Svenska English

Start page
Collections
Collaborating Partners
Frequently asked questions
About Naturarv

Advanced search [Close advance search](#)

	Contains	Classification	<input type="text"/>	-
	AND	Contains	Swedish name	-
	OR	Contains	Type status	-
			<input type="checkbox"/> Is type	-
	NOT	Contains	Determined by	- +

Figure 15. Interface that supports Boolean operators at the Swedish Museum of Natural History

- (3) Gustavianum (<https://pragmata.sia.uu.se/pragmata/>), which requires the user to type "+" to give a search term higher priority (this is the phrase used in the interface), "-" for NOT, and quotation marks for an exact phrase.
- (4) Carlotta gives no specific instructions; the help page on searching says only that terms can be combined with each other. An exception is made for the use of a wild card ("*"), the usage of which is explained. However, upon testing a few examples, it seems that the following is supported: AND, OR, NOT, "?", "*" and quotation marks.
- (5) MIMO is similar: some Boolean operators work ("AND", "NOT", "?" and "*") but no help is provided on how to search.

- (6) DigitaltMuseum, eMuseumPlus and Europeana all support the use of “AND”, “OR” and “NOT”. In eMuseumPlus and Europeana truncation is available, and in Europeana the use of a wild card (“?”) and a similar spelling function (“~”) are also available.
- (7) Wikimedia Commons provides a more user-friendly support for the use of Boolean operators in its advanced search interface, shown in [Figure 16](#) below: “These words” (AND), “Not these words” (NOT) and “One of these words” (OR), as well as searching by phrase with “Exactly this text”.

Feature 14, linking each subject access point to its resources, is a key feature of hypertext that ought to be fully explored in online search services. There are 11 museums with their own search tools (Sörmland Museum, the Museum of Work, 2 of 2 Digitala Stadsmuseet museums and 7 of 7 Kulturhotell museums) and only 1 cross-search service (Carlotta) whose interface supports this functionality, a total of 25 museums (27.5%). All of these work in the similar way: the user can click on a keyword in the metadata record of the retrieved object in order to retrieve other objects matching the same keyword. Carlotta supports this feature in many metadata fields, of which only two are subject related: what kind of object is represented and the place shown on the image.

4.2.6 *Features 15 and 16.* Feature 15, linking subject access points from one controlled vocabulary to corresponding concepts in others, would allow interoperability and enhanced subject access enriched by related concepts from other vocabularies; however, this has not been implemented in any of the services.

Feature 16, adding, browsing and searching end user tags, would be useful to provide end-user perspectives on topics represented by museum objects, but has not been implemented by any services. The Swedish History Museum appears to have a beta version underway. At the time of this research, the tags are not social media tags but instead represent only a ranking of the most frequently used search words from their vocabularies. The users can also enter social tags, but they cannot search the museum objects. DigitaltMuseum claims that it gives logged-in users an option to add tags, but no tag clouds or related functionalities could be found.

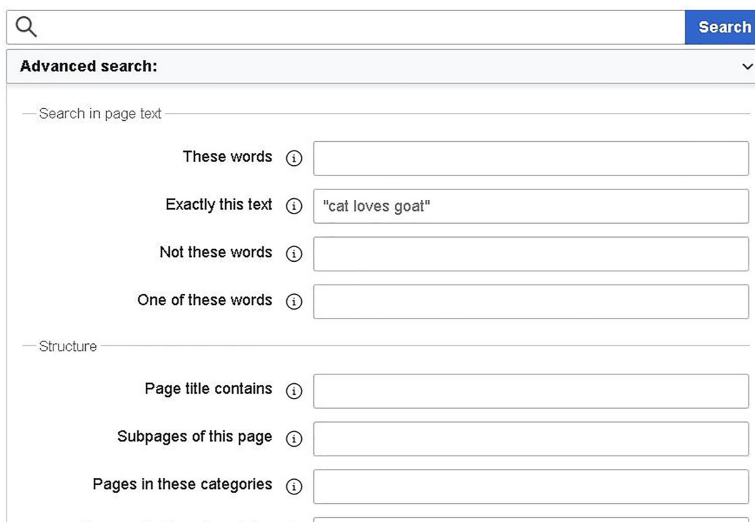


Figure 16.
A more user-friendly way to implement Boolean operators and phrase searching, implemented by Wikimedia Commons

4.2.7 *Features 17 and 18.* Feature 17, combining previous search formulations, is important for advanced search queries and in saving time for the end user. Six museums have their own search tools (the Swedish Museum of Natural History, Sörmland Museum, the Museum of Natural History and 3 individual eMuseumPlus museums), and four cross-search services (Alvin, Carlotta, Digitalt Museum and eMuseumPlus) use the feature, a total of 63 museums (69.2%). The feature is implemented similarly in all of them: when the user makes any subsequent search, the previous search phrase is still present in the search box and can be used again.

Feature 18, help on searching, is provided by 16 museums' in-house search tools and 6 cross-search services, a total of 80 museums (87.9%). The feature is implemented in two general ways. The most common is a dedicated web page with all the necessary instructions (the Swedish Museum of Performing Arts, Stockholm County Museum, Gustavianum, Alvin, Carlotta, DigitaltMuseum, eMuseumPlus, Europeana).

A more user-friendly approach is to provide contextual help: for example, by providing question mark icons at different places in the interface (the Swedish History Museum, Sörmland Museum, the Paleontological Museum of Uppsala University, Kulturhotell, Wikimedia Commons). [Figure 17](#) below shows an example of the Paleontological Museum of Uppsala University.

4.2.8 *Features 19, 20, 21.* These three features all relate to images, which are a key characteristic of museum collections and should be offered at full scale. Feature 19, searching by image-related features (e.g. adding search criteria to retrieve results/records with or without corresponding images; image orientation, i.e. portrait or landscape, image size) is provided by six museums' in-house search tools (the Swedish Museum of Natural History, Moderna Museet and 3 individual eMuseumPlus museums), and four cross-search services (Carlotta, eMuseumPlus, Europeana and Kringla) use this feature (72 museums or 79.1%). All

Search the database

UPS-THUNB No. ? Exact

UPS catalogue No. ?

Organism group Vascular plants Bryophytes Lichens
 Algae Fungi All

Original name ? Exact

Current name Exact

Family

family_ja

Identification history ?

Locality

Notes on the sheet ?

Type status

Basionym ? Exact

Nomenclatural notes ?

Icones Exact

Figure 17. Contextualized help at the Paleontologic Museum of Uppsala University

of these search tools except Europeana let the user choose to retrieve only results with images. Europeana does not have this option; however, it lets the user choose the image orientation (portrait or landscape) and image size.

Feature 20, searching by Content-based Image Retrieval (CBIR) methods (e.g. query by example image; query by sketch map; query by colour map) is supported by two cross-search tools: Europeana and Google Arts and Culture, that is, 60 museums (65.9%). Both services allow searching by colour. [Figure 1](#) (in the Background section) shows an example of searching by colour in the Europeana interface.

Feature 21, encompassing functionalities encompassed to implementation of the International Image Interoperability Framework (IIIF) open standards for enhanced user experience (e.g. deep zoom viewing, comparing, manipulating and annotating images, API interfaces), is applied in two cross-search services (4 museums or 4.4%). These support only deep zoom viewing: Google Arts and Culture after clicking the magnifier icon by the picture; and Wikimedia Commons, when choosing “Open in Media Viewer”.

4.3 Summary

Based on the analysis of all the features, Sörmland Museum and Kulturhotell are the search interfaces that use significantly more search features than other interfaces. On the opposite end of the spectrum lie Västernorrlands Museum, the Museum of Sketches for Public Art, Arlanda flygsamlingar (Arlanda Civil Aviation Collection), Waldemarsudde, Teleseum and the Thiel Gallery which do not support any search feature from our list. Implications for users may vary; for example, while it is much easier to find, identify, select, obtain and explore the documents that match the information need when using Sörmland Museum and Kulturhotell than in those using few or none of the features, because of the very specific and limited character of some collections their rudimentary search interfaces may suffice in some contexts. The latter may be the case in Arlanda flygsamlingar, Waldemarsudde and Teleseum where short lists of objects reflect the museums’ collections, but this is much less obvious in other museums such as the Västernorrlands Museum. Finally, it would be useful to know whether the limitations ensue from the interface *per se* or the metadata and related indexing policies; this should be studied in the future by applying another method such as interviews with interface and metadata creators.

5. Conclusion

Subject searching is a very common, yet the most complex, type of search carried out by end users in online information services provided by cultural heritage institutions. Although many standards, guidelines and practices are in place to this effect, quality-controlled subject indexing and appropriate information retrieval interfaces which take advantage of this indexing seem to be largely missing from online search services at Swedish museums. This matches the findings with regard to databases of journal articles ([Golub et al., 2020](#)), repositories ([Golub et al., 2020](#)) and discovery services ([Golub, 2018](#)). This study assessed the websites of 91 museums, all of which were found to provide online access to at least some of their holdings, and 9 cross-search services. The study analysed the search interfaces against a set of 21 criteria and showed that effective subject access is largely unavailable in existing services. Few of these support hierarchical browsing of classification schemes and other controlled vocabularies with hierarchical structures, few provide end-user-friendly options to choose a more specific concept to increase search precision, suggest a broader concept or related concepts to increase recall, disambiguate homonyms, or find which term is best to name a particular concept.

In fact, we have not found a single confirmed case of an established subject-related controlled vocabulary in these services. This also makes cross-searching across combined databases very challenging, since there is no such control within individual databases, let alone

any mapping between vocabularies across the databases. While it has been determined previously that inconsistent and incomplete metadata and blending of controlled vocabularies, free keywords and full-text automatic indexing create the biggest problems for subject searching (Dempsey, 2012; Fagan, 2011), here the situation is further exacerbated by the fact that no controlled vocabularies are used whatsoever. There seem to be efforts under way to alleviate this: KulturNav (<https://kulturnav.org>) is envisioned as a platform for creating, managing and distributing linked open name authorities and vocabularies for cultural heritage.

There is a strong need for the implementation of established controlled vocabularies in museums more widely, not only in Sweden. The heterogeneity of object types and the uniqueness of museum materials are a factor in the underuse and even underdevelopment of terminology for the techniques, types and functions of these objects and consequently for their subjects. Even the AAT, the most comprehensive thesaurus for the cultural heritage domain, is constantly evolving through the addition of new concepts. The AAT is multilingual, and translation projects into many languages are currently active: so the need to translate concepts and definitions into Swedish should be emphasized in particular here. In addition, it is important to record unique local terminology in ethnographic museums and museums of local communities more widely, which goes beyond the scope of the AAT.

Access to image-based resources is fundamental both to research and to the transmission of cultural heritage knowledge; therefore, in order to enable reliable recognition and interpretation of a subject, users need to be provided with high-quality zoomable photographs in which subjects (e.g. objects, motifs, persons, etc.) should be visually indicated and/or annotated in an interoperable way according to the IIF standard. Computer vision is an important technology which could help museums to recognize real world objects in images (i.e. which can significantly improve progress in indexing large collections of photographs), based on research in deep learning as a part of the field of artificial intelligence. It is important to note that applying this kind of automatic indexing in end-user search interfaces requires informing the user transparently about where the subject index terms come from, whether assigned by humans (experts or end users) or by machines.

This article also identifies some functionalities which do not even exist in the online catalogues of the world's most prestigious museums but are important for the correct interpretation of a subject under analysis. Due to the limitations and biases caused by the cultural conditionality and subjectivity of subject analysis, indexing and interpretation, it is important to provide users with an insight into the sources of knowledge according to which type of subject analysis was carried out at the interface level (e.g. related scholarly sources, textual and audio-visual documentation, field research results, witnessing information and other sources of knowledge). Users should also be informed about the level of subject analysis: *description, identification, interpretation, isness, aboutness, ofness*. The development of a new version of the LIDO data exchange standard should facilitate the display of this functionality in the user interface.

Future research should focus specifically on user interfaces for subject access, how best to support query expansion, word sense disambiguation, etc., based on specific user needs. All of these refinements should be securely based in user studies, analysis of real search sessions including all potential user groups such as humanities scholars, interdisciplinary scholars, university students, cultural heritage professionals and the general public.

This study should also be complemented by future research on subject access in archives. Furthermore, comparative studies across international borders would help determine the status of this area of research elsewhere, as it is likely that the situation is similar and would require regional, European or worldwide policies to be updated and/or resources to be put in place to improve this situation.

While the literature points to some reasons for this, this study should be complemented in future by a qualitative study, an interview (individual or focused group) in order to determine

the reasons for the current poor state of subject access in cultural heritage institutions, and a Delphi study targeting researchers who are experts in this field, with policymakers and other stakeholders or decision-makers in order to reach more informed decisions about policies and distribution of resources to improve this situation. While it is highly likely that much of this lack of development can be attributed to scarce resources which are increasingly stretched to cover an ever-growing number of tasks in cultural heritage institutions, the authors call upon ICOM-CIDOC, IFLA and ICA (the International Council of Archives) to act together to create subject access guidelines for all cultural heritage institutions. Finally, as we see that museums are already collaborating with Google and Wikimedia, and at times even outsourcing their collections for user access to them, companies which are not directly involved with CIDOC or related professional organizations, it is important that those services become more familiar with professional advice from information retrieval specialists in order to meet the needs of users.

References

- Alani, H., Jones, C. and Tudhope, D. (2000), "Associative and spatial relationships in thesaurus-based retrieval", in Borbinha, J. and Baker, T. (Eds), *Proceedings (ECDL 2000) 4th European Conference on Research and Advanced Technology for Digital Libraries*, Lecture Notes in Computer Science, Springer, Berlin, pp. 45-58. doi: [10.1007/3-540-45268-0_5](https://doi.org/10.1007/3-540-45268-0_5).
- Alvin (2021), "Alvin:Info", available at: <https://info.alvin-portal.org/> (accessed 20 February 2021).
- Baca, M. (2004), "Fear of authority?: authority control and thesaurus building for art and material culture information", *Cataloging and Classification Quarterly*, Vol. 38 Nos 3-4, pp. 143-151, doi: [10.1300/J104v38n03_13](https://doi.org/10.1300/J104v38n03_13).
- Baca, M. and Harpring, P. (2016), *Categories for the Description of Works of Art (CDWA)*, Getty Research Institute, available at: https://www.getty.edu/research/publications/electronic_publications/cdwa/ (accessed 5 April 2021).
- Baca, M., Harpring, P., Lanzi, E., McRae, L. and Whiteside, A. (Eds) (2006), *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*, American Library Association, available at: <http://vraweb.org/wp-content/uploads/2018/08/CatalogingCulturalObjectsFull.pdf> (accessed 2 April 2021).
- Bair, S. and Carlson, S. (2008), "Where keywords fail: using metadata to facilitate digital humanities scholarship", *Journal of Library Metadata*, Vol. 8 No. 3, pp. 249-262, doi: [10.1080/19386380802398503](https://doi.org/10.1080/19386380802398503).
- Bates, M.J. (1996), "The Getty end-user online searching project in the humanities: report no. 6: overview and conclusions", *College and Research Libraries*, Vol. 57 No. 6, pp. 514-523, doi: [10.5860/crl_57_06_514](https://doi.org/10.5860/crl_57_06_514).
- Caplan, P. (1995), "You call it corn, we call it syntax-independent metadata for document-like objects", *The Public-Access Computer Systems Review*, Vol. 6 No. 4, pp. 19-23.
- Carboni, N. and Luca, L.D. (2017), "Towards a semantic documentation of heritage objects through visual and iconographical representations", *International Information and Library Review*, Vol. 49 No. 3, pp. 207-217, doi: [10.1080/10572317.2017.1353374](https://doi.org/10.1080/10572317.2017.1353374).
- Coburn, E., Light, R., McKenna, G., Stein, R. and Vitzthum, A. (2010), *LIDO-Lightweight Information Describing Objects Version 1.0*, ICOM International Committee of Museums, available at: <http://lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>.
- Cutter, C.A. (1876), *Rules for a Printed Dictionary Catalogue*, Government Printing Office, Washington.
- Dempsey, L. (2012), "Thirteen ways of looking at libraries, discovery, and the catalog: scale, workflow, attention", *EDUCAUSE Review*, available at: <https://er.educause.edu/articles/2012/12/thirteen-ways-of-looking-at-libraries-discovery-and-the-catalog-scale-workflow-attention> (accessed 29 November 2019).

- DigitaltMuseum (2021), "About DigitaltMuseum", available at: <https://dok.digitaltmuseum.org/en/about> (accessed 20 February 2021).
- Dobrev, M. and Chowdhury, S. (2010), "A user-centric evaluation of the Europeana digital library", *International Conference on Asian Digital Libraries*, June, Springer, Berlin and Heidelberg, pp. 148-157.
- Doerr, M., Gradmann, S., LeBoeuf, P., Aalberg, T., Bailly, R. and Olensky, M. (2013), *Final Report on EDM-FRBRoo Application Profile Task Force, Technical Report, Europeana Professional*, p. 33, available at: <http://www.cidoc-crm.org/frbroo/sites/default/files/TaskfoApplication%2BProfile%2BEDM-FRBRoo.pdf> (accessed 2 April 2021).
- eMuseumPlus (2021a), "Hjälp att söka", available at: <http://emuseumplus.lsh.se/eMuseumPlus?service=WebAsset&url=html/searchHelp.html&contentType=text/html> (accessed 20 February 2021).
- eMuseumPlus (2021b), "Om samlingarna och databasen", available at: <http://emuseumplus.lsh.se/eMuseumPlus?service=WebAsset&url=html/aboutUs.html&contentType=text/html> (accessed 20 February 2021).
- Europeana Foundation (2017), "Europeana data model – mapping guidelines v2.4", available at: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf.
- Europeana Foundation (2021), "About Europeana", available at: <https://www.europeana.eu/en/about-us> (accessed 20 February 2021).
- Fagan, J.C. (2011), "Discovery tools and information literacy", *Journal of Web Librarianship*, Vol. 5 No. 3, pp. 171-178, doi: [10.1080/19322909.2011.598332](https://doi.org/10.1080/19322909.2011.598332).
- Fantoni, S.F., Stein, R. and Bowman, G. (2012), "Exploring the relationship between visitor motivation and engagement in online museum audiences", *Museums and the Web*, available at: https://www.museumsandtheweb.com/mw2012/papers/exploring_the_relationship_between_visitor_mot.
- Fortier, A. and Ménard, E. (2018), "What do museum website users expect from linked open data?", *Advances in Knowledge Organization*, Vol. 16, pp. 900-907, available at: <http://ocs.lettras.up.pt/index.php/fisko2018/isko2018/paper/view/1445>.
- Freire, N., Robson, G., Howard, J.B., Manguinhas, H. and Isaac, A. (2020), "Cultural heritage metadata aggregation using web technologies: IIF, Sitemaps and Schema.org", *International Journal on Digital Libraries*, Vol. 21 No. 1, pp. 19-30, doi: [10.1007/s00799-018-0259-5](https://doi.org/10.1007/s00799-018-0259-5).
- Gaona-García, P., Feroso, A. and Sánchez, S. (2017), "Exploring the relevance of Europeana digital resources: preliminary ideas on Europeana metadata quality", *Revista Interamericana De Bibliotecología*, Vol. 40 No. 1, pp. 59-69, doi: [10.17533/udea.rib.v40n1a06](https://doi.org/10.17533/udea.rib.v40n1a06).
- Golub, K. (2003), "Predmetno pretraživanje u knjižničnim katalozima s web- sučeljem", [Subject searching in web-based library catalogs] (Unpublished master's thesis), University of Zagreb, Zagreb, available at: <http://koraljka.info/publ/Magisterij-hrv.pdf>.
- Golub, K. (2016), "Potential and challenges of subject access in libraries today on the example of Swedish libraries", *International Information and Library Review*, Vol. 48 No. 3, pp. 204-210, doi: [10.1080/10572317.2016.1205406](https://doi.org/10.1080/10572317.2016.1205406).
- Golub, K. (2018), "Subject access in Swedish discovery services", *Knowledge Organization*, Vol. 45 No. 4, pp. 297-309, doi: [10.5771/0943-7444-2018-4-297](https://doi.org/10.5771/0943-7444-2018-4-297).
- Golub, K., Tyrkkö, J., Hansson, J. and Ahlström, I. (2020), "Subject indexing in humanities: a comparison between a local university repository and an international bibliographic service", *Journal of Documentation*, Vol. 76 No. 6, pp. 1193-1214, doi: [10.1108/JD-12-2019-0231](https://doi.org/10.1108/JD-12-2019-0231).
- Google (2021), "Bringing the world's art and culture online for everyone", available at: <https://about.artsandculture.google.com/> (accessed 20 February 2021).
- Harpring, P. (2018), "Linking the getty vocabularies: the content perspective, including an update on CONA", *2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*. Presented at the 2018 Pacific Neighborhood Consortium Annual Conference

and Joint Meetings (PNC), IEEE, San Francisco, California, pp. 1-8. doi: [10.23919/PNC.2018.8579460](https://doi.org/10.23919/PNC.2018.8579460).

Swedish online
museums

Heery, R., Lyon, L., Tsinaraki, C., Brody, T., Koch, T. and Doerr, M. (2006), *Report on Digital Repositories: An Evaluation Study on the Development and Implementation of Community Repositories to Support Research (And Learning and Teaching)*, DELOS2 Network of Excellence on Digital Libraries, Deliverable 5.1.1, available at: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=19458833E5FA5CC45117776085BD0E87?doi=10.1.1.101.5976&rep=rep1&type=pdf> (accessed 30 November 2019).

Hider, P. and Liu, Y.-H. (2013), "The use of RDA elements in support of FRBR user tasks", *Cataloging and Classification Quarterly*, Vol. 51 No. 8, pp. 857-872, doi: [10.1080/01639374.2013.825827](https://doi.org/10.1080/01639374.2013.825827).

Hunter, N.R. (1991), "Successes and failures of patrons searching the online catalog at a large academic library: a transaction log analysis", *RQ*, Vol. 30 No. 3, pp. 395-402, available at: <https://www.jstor.org/stable/25828813> (accessed 30 November 2019).

ICOM and CIDOC (2002), "Issue 14: how to model 'subjects'", available at: <http://www.cidoc-crm.org/Issue/ID-14-how-to-model-subjects>.

ICOM and CIDOC (2006), available at: <http://www.cidoc-crm.org/frbroo/home-0>.

ICOM Statutes (2007), "Museum definition", available at: http://icomsweden.se/wp-content/uploads/2010/12/statutes_eng.pdf.

Iconclass (2021), "RKD Nederlands Instituut voor Kunstgeschiedenis", available at: <https://rkd.nl/nl/collecties/services-tools/iconclass> (accessed 5 April 2021).

IIIF – International Image Interoperability Framework (2021), available at: <https://iiif.io/> (accessed 2 April 2021).

ISO (1985), *Documentation – Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms (5963:1985)*, available at: <https://www.iso.org/standard/12158.html>.

ISO (2006), *Information and Documentation – A Reference Ontology for the Interchange of Cultural Heritage Information (21127:2006)*, available at: <https://www.iso.org/standard/34424.html>.

Jack, C. (2001), "State of the arts: current applications for indexing images", available at: <https://web.archive.org/web/20010210080529/http://www.slis.ualberta.ca/599/cjack/599.htm> (accessed 9 January 2021).

Knapp, S.D., Cohen, L.B. and Juedes, D.R. (1998), "A natural language thesaurus for the humanities: the need for a database search aid", *Library Quarterly*, Vol. 68 No. 4, pp. 406-430, doi: [10.1086/603001](https://doi.org/10.1086/603001).

Koutsomitropoulos, D.A., Hyvönen, E. and Papatheodorou, T.S. (2012), "Editorial: semantic web and reasoning for cultural heritage and digital libraries", *Semantic Web – Interoperability, Usability, Applicability*, Vol. 3 No. 1, p. 1.

Kuhagen, J.A. (2015), "Subject relationship element in RDA chapter 23", available at: <http://www.rda-jsc.org/archivedsite/docs/6JSC-ALA-31-rev-Sec-final.pdf>.

Kulturen (2021), "Carlotta – ett databssystem för museisamlingar", available at: <http://carl.kulturen.com/web> (accessed 20 February 2021).

Liew, C.L. (2004), "Online cultural heritage exhibitions: a survey of information retrieval features", *Program Electronic Library and Information Systems*, Vol. 39 No. 1, pp. 4-24, doi: [10.1108/00330330510578778](https://doi.org/10.1108/00330330510578778).

Markey, K. (2007), "The online library catalogue: paradise lost and paradise regained?", *D-Lib Magazine*, Vol. 13 Nos 1/2, doi: [10.1045/january2007-markey](https://doi.org/10.1045/january2007-markey).

Meadow, K. and Meadow, J. (2012), "Search query quality and web-scale discovery: a qualitative and quantitative analysis", *College and Undergraduate Libraries*, Vol. 19 Nos 2-4, pp. 163-175, doi: [10.1080/10691316.2012.693434](https://doi.org/10.1080/10691316.2012.693434).

MIMO (2021), "About MIMO", available at: <https://mimo-international.com/MIMO/about-mimo.aspx> (accessed 20 February 2021).

Panofsky, E. (1993), *Meaning in the Visual Arts*, Penguin Group, Harmondsworth.

- Patel, M., Koch, T., Doerr, M. and Tsinaraki, C. (2005), *Semantic Interoperability in Digital Library Systems*, DELOS2 Network of Excellence on Digital Libraries, Deliverable 5.3.1, available at: <http://delos-wp5.ukoln.ac.uk/project-outcomes/SI-in-DLs/SI-in-DLs.pdf> (accessed 30 November 2019).
- RDA Co-Publishers (2017), "RDA toolkit: resource description and access", available at: <https://access.rdatoolkit.org>.
- Riksantikvarieämbetet (2019), "Om Kringla", Updated May 3, 2019, available at: <https://www.raa.se/hitta-information/kringla/om-kringla/>.
- Riva, P., Le Bœuf, P. and Žumer, M. (2017), "IFLA library reference model: a conceptual model for bibliographic information", *International Federation of Library Associations and Institutions*, August.
- Seadle, M. (2010), "Archiving in the networked world: interoperability", *Library Hi Tech*, Vol. 28, pp. 189-194, doi: [10.1108/07378831011047604](https://doi.org/10.1108/07378831011047604).
- Siegfried, S., Bates, M.J. and Wilde, D.N. (1993), "A profile of end-user searching behavior by humanities scholars: the Getty online project report no. 2", *Journal of the American Society for Information Science*, Vol. 44 No. 5, pp. 273-291, doi: [10.1002/\(SICI\)1097-4571\(199306\)44:53.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-4571(199306)44:53.0.CO;2-X).
- Skov, M. and Ingwersen, P. (2014). "Museum web search behaviour of special interest visitors", *Library and Information Science Research*. Vol. 36 No. 2, pp. 91-98, doi: [10.1016/j.lisr.2013.11.004](https://doi.org/10.1016/j.lisr.2013.11.004).
- Social History and Industrial Classification (2021), available at: <http://www.shcg.org.uk/About-SHIC> (accessed 17 March 2021).
- Srinivasan, R., Boast, R., Becwar, K.M. and Furner, J. (2009), "Blobgects: digital museum catalogs and diverse user communities", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 4, pp. 666-678.
- Svenonius, E. (1994), "Access to nonbook materials: the limits of subject indexing for visual and aural languages", *Journal of the American Society for Information Science*, Vol. 45 No. 8, pp. 600-606.
- Swedish National Heritage Board (2021), "About SOCH", available at: <https://www.raa.se/in-english/digital-services/about-soch/> (accessed 8 April 2021).
- The Getty Foundation (2012), *Moving Museum Catalogues ONLINE*, An Interim Report from the Getty Foundation, available at: https://www.getty.edu/foundation/pdfs/osci_interimreport_2012.pdf.
- Tibbo, H.R. (1994), "Indexing for the humanities", *Journal of the American Society for Information Science*, Vol. 45 No. 8, pp. 607-619, doi: [10.1002/\(SICI\)1097-4571\(199409\)45:83.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-4571(199409)45:83.0.CO;2-X).
- Trant, J. (2009a), "Studying social tagging and folksonomy: a review and framework", *Journal of Digital Information*, Vol. 10 No. 1, available at: <http://hdl.handle.net/10150/105375>.
- Trant, J. (2009b), "Tagging, folksonomy and art museums: results of steve.museum's research 2009-01", available at: <https://repository.arizona.edu/bitstream/handle/10150/106510/trant-taggingArt.pdf?sequence=1>.
- Trant, J., Wyman, B. and with the Participants in the steve.museum Project (2006a), "Investigating social tagging and folksonomy in art museums with steve.museum", available at: https://www.researchgate.net/publication/240730620_Investigating_social_tagging_and_folksonomy_in_art_museums_with_SteveMuseum.
- Trant, J. and with the Participants in the steve.museum Project (2006b), "Exploring the potential for social tagging and folksonomy in art museums: proof of concept", *New Review of Hypermedia and Multimedia*, Vol. 12 No. 1, pp. 83-105, doi: [10.1080/13614560600802940](https://doi.org/10.1080/13614560600802940).
- Tudhope, D., Binding, C., Blocks, D. and Cunliffe, D. (2006), "Query expansion via conceptual distance in thesaurus indexed collections", *Journal of Documentation*, Vol. 62 No. 4, pp. 509-533, doi: [10.1108/00220410610673873](https://doi.org/10.1108/00220410610673873).
- Villaespesa, A. (2017), "Who are the users of the met's online collection?", available at: <https://www.metmuseum.org/blogs/collection-insights/2017/online-collection-user-research> (accessed 8 April 2021).

-
- Villaespesa, E., Tate, U. and Stack, J. (2015), "Finding the motivation behind a click: definition and implementation of a website audience segmentation", in *MW2015: Museums and the Web 2015*, available at: <https://mw2015.museumsandtheweb.com/paper/finding-the-motivation-behind-a-click-definition-and-implementation-of-a-website-audience-segmentation/>.
- Villén-Rueda, L., Senso, J.A. and De Moya-Anegón, F. (2007), "The use of OPAC in a large academic library: a transactional log analysis study of subject searching", *The Journal of Academic Librarianship*, Vol. 33 No. 3, pp. 327-337.
- Waibel, G., LeVan, R. and Washburn, B. (2010), "Museum data exchange: learning how to share", *D-Lib Magazine*, Vol. 16 Nos 3/4, doi: [10.1045/march2010-waibel](https://doi.org/10.1045/march2010-waibel).
- Wallis, R., Isaac, A., Charles, V. and Manguinhas, H. (2017), "Recommendations for the application of Schema.org to aggregated Cultural Heritage metadata to increase relevance and visibility to search engines: the case of Europeana", *Code4Lib Journal*, Vol. 36, p. 1, available at: <https://journal.code4lib.org/articles/12330>.
- Walsh, D., Clough, P. and Foster, J. (2016), "User categories for digital cultural heritage", *First International Workshop on Accessing Cultural Heritage at Scale*, pp. 3-9, available at: https://www.researchgate.net/publication/304114334_User_Categories_for_Digital_Cultural_Heritage.
- Walsh, D., Hall, M.H., Clough, P. and Foster, J. (2018), "Characterising online museum users: a study of the National Museums Liverpool Museum website", *International Journal on Digital Libraries*, Vol. 21, pp. 75-87, doi: [10.1007/s00799-018-0248-8](https://doi.org/10.1007/s00799-018-0248-8).
- Wikimedia Commons (2021), "What is Wikimedia Commons?", Updated January 4, 2021, available at: <https://commons.wikimedia.org/wiki/Commons:Welcome>.
- Wikipedia (2020), "List of museums in Sweden", available at: https://en.wikipedia.org/wiki/List_of_museums_in_Sweden (accessed 1 July 2020).
- Will, L. (1993), "The indexing of museum objects", available at: https://www.theindexer.org/files/18-3/18-3_157.pdf.
- Zeng, M., Žumer, M. and Salaba, A. (Eds) (2011), in *Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model*, De Gruyter Saur, Berlin and New York.
- Zhou, W., Li, H. and Tian, Q. (2017), "Recent advance in content-based image retrieval: a literature survey", arXiv:1706.06064 [cs], available at: <http://arxiv.org/abs/1706.06064> (accessed 30 January 2021).

Corresponding author

Koraljka Golub can be contacted at: koraljka.golub@lnu.se

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com