Ari Häyrinen
PhD-student, Digital Culture
Department of Art and Culture Studies
University of Jyväskylä
Juomatehdas, Vapaudenkatu 25
tel. (014) 260 1460
fax. (014) 260 1461
email: arihayri@jyu.fi

Google, world's largest museum?

abstract

Studies and everyday practise have shown that generic search engines - Google, for example - are the most important entry points to the Internet. Therefore it is surprising to see how weak the presence of museums in the search engine universe is. Unless the search query includes the name of the museum, it is very likely that the top hits for cultural historical queries are not museum sites – instead, first to appear are individual blog articles, Wikipedia, different kinds of volunteer-based virtual museums, commercial sites and occasional home pages. Secondly, even if the resource provided by the museum is found, the content is often traditional collection management data that is not designed for public use and looks very sparse and non-contextualised when compared to the information found from other sources.

I will analyze the current situation and reasons for the digital silence of museums in the search engine universe. The question is whether this silence is a problem or not, and if it is a problem, then what kind of actions could be taken in order to improve the situation knowing the limited resources museums have. I argue that what is needed is a knowledge management model accepting the fact that very relevant information sources on museum collections lies outside museums and that these resources should be seen as a possibility, not as a threat. This model relies on sharing visual content, crowdsourcing and using unique images as linking elements between sources. The model is demonstrated with example cases.

Table of Contents

Introduction....1
Current situation by examples....4
James Bond’s toy gun....4
National costumes of Finland....5
A cultural environmental portal....6
Bowden Spacelander....7
Mărețele maimuțe and William Holbrook Beard....8
Goya's paintings....8
American Bison at the Castle....9
Discussion....9
The importance of images....10
Sharing or just displaying?....10
Solutions....12
Conclusion....12

Introduction

“To Google” is a verb meaning the use of Google search engine to find things from the World Wide Web. According to OCLC, in 2010 approximately 84 per cent of American online users started their information search with a search engine (OCLC, 2010). Google-type searches were the most frequent way for finding resources for a faculty of humanities and social sciences (Harley, 2007). Students use Google so much that some teachers have even banned its use in their classes (Chiles, 2008). These examples confirm the fact most users, including average citizens, students and even teachers use Google to find information resources.

A great amount of institutional cultural heritage material can be found online. In addition, there are independent non-institutional online resources competing with museums, archives and libraries. Instead of just overlapping with the sources provided by the traditional institutions, these novel sources often surpass the institutional sources in search results and coverage. The question now is: how will museums react? Do they want to be part of that information network or do they want to keep aloof in order to maintain their authority and the control over their materials?

As far back as one can remember, there have been researchers and enthusiastic amateurs who have written and published their own works independently, without being supported by institutions. What has changed in our digital age is the ease of creating and finding digital resources. Everyone can contribute to the pool of digital heritage since there are no gatekeepers in the open Internet. No matter how poor or good the information contributed, it is picked up by search engines. This puts all heritage institutions to a new situation in the open world of Internet: they have lost their monopoly to “say how things really are” as well as their automatic top ratings in search results. In the search engine universe the institutional status of the heritage organisations disappears, and they are presented on the same level as any other resources.

Before examining these questions more closely, we must make a distinction between the web presence of the museum as such and the presence of the museum in the web as an information resource visible to search engines. This distinction is clear when we think about two contrasting web search situations: in the first one, a person is planning to visit a museum and, in the second one, a web user is searching information about a certain topic or item. In the first case, the person wants to know what kinds of exhibitions the museums have, view their opening hours and possibly browse their collections. What is presented in the first case is the web interface that is strictly controlled by the museum. The second case shows the backside of the web interface, i.e. what is seen by the search engines. In the search engine universe the content provided by the museum is on the same level as any other source of information. I will shortly turn the focus on museums' web presence level.

In this paper I will study how cultural heritage information and materials are divided between heritage organisations and unofficial sources in the search engine universe as well as the reasons behind this phenomena. I have chosen examples that demonstrate different kinds of problems, solutions and attitudes of heritage organisations. I also study the role of images, sharing of images and image based search as a possible link between institutional and free online resources.

Previous studies

There seems to be two assumptions in museum literature and museum informatics literature related to museums' online presence. The first concerns the role of the museum: the museum is - and it should always be - an active participant. It is the museum that defines the framework for all activities that

online users are able to do online, leaving very little room for the users' own innovations. The museum dominates the process and makes sure that everything is properly authorised, that the facts are right and that inappropriate behaviour won't occur. The second assumption is that the user has already entered the context of the museum's online facilities. The following example illustrates this assumption:

> *A student writing a paper on the "Labors of Herakles", for instance, might expect to be able to ask to see all object in the museum's collection that are related to that topic. (Marty, 2008)*

The question remaining is: how did the student find the museum's collection in the first place? In most of the studies there is an unexplained gap between web users' search behaviour and the starting point for their search which comes across a museum online. Users starting their search with Google and finding no references to museum collections will rarely come across the museums' online resources. In addition, in order to be found the resource should have quite a high ranking in the results since often only the first two search pages are browsed through by the user.

> *They [students] reported that they often find what they need within the first two pages of results and rarely feel the need to view more than what is shown on the first two screens.(Rieger, 2009)*

Leif Isaksen and Sebastien Heath have both pointed out that online information about cultural heritage objects is not found only on traditional memory organisations' websites. Quite to the contrary, in some cases information is found mainly outside the "official" resources. In his article Pandora's box, Leif Isaksen uses Mona Lisa as an example. Isaksen points out that search with Google lists first Wikipedia, then YouTube and some other sites before Louvre's page about Mona Lisa appears. Isaksen claims that digital isolation of heritage institutions cannot be justified and that it certainly does not improve their current situation.

> *Let us repeat, then, that it is no longer worth arguing whether we should try to control cultural information. Clinging on to artefacts and archives, restricting or obscuring data, claiming academic authority – these are of little or no use to us if we are struggling even to make ourselves heard.*
>
> *[...] we need to accept that an isolationist policy amongst heritage professionals is not only foolish but ultimately meaningless because culture will carry on without us. (Isaksen, 2009)*

Although Isaksen can be criticised only having one - and very famous - example, this example demonstrates the current situation. Information about heritage is divided to several sources and heritage institutions are not necessarily at the top of the list.

> *We have to accept that not only do we not know everything, but that cultural perspectives are infinite in principle. Nonetheless, there are plenty of tools available that allow us to monitor and moderate input from 'the crowd' to our own sites. (Isaksen, 2009)*

Instead of thinking of heritage institutions as a source of authoritative information (the "high culture" as Isaksen states) that web users can see, we could define these institutions as content holders that seek users for their content. This is a really a radical change of view, and it as much counter to the current situation as possible. Isaksen entertains a thought about a certain kind of cultural relativism although he does not use that term. An infinite amount of cultural perspectives means that there is no one perspective that should dominate.

Sebastian Heat discusses digitization in numismatic and Roman pottery, resulting from the commercial sale of antiquities. In the field of ancient numismatic, 80 percent of the sites found by Google were commercial or personal sites. Although many of the online resources does meet the best practises in the

fields in question, it does not mean that these resources should be excluded (Heath, 2010). Heat's examples show that there are online resources that surpass, both in quality and accessibility, the resources provided by heritage institutions. There are great variations in the quality of the sources, and different fields have varying levels of activity. The activity generated by the commercial sale of antiquities produces large amounts of information, and academic institutions find it difficult to compete with that activity.

Melissa Terras has studied amateur digitisation resources. She concentrates on online resources that are run by keen individuals who wish to participate in digitizing cultural heritage. To start with, she selected one hundred resources, and then she interviewed six of the creators of high-quality resources. She summarises the importance of these resources as follows:

> *[...] that most presented novel, detailed, and niche content with a very specific scope. Ephemera which had not been collected—or even noticed—elsewhere was documented, stored, presented, and catalogued. (Terras, 2010)*

The best resources are actively maintained; they are widely used – also in academic fields –; and they often appear near the top of the listing of topical hits by a search engine.

> *Additionally, those creating such online materials are generally more successful in interacting with their relevant online communities than memory institutions are. As a result, instead of being viewed as mere digital 'cabinets of curiosities', the best digital resources created by enthusiasts, in their own time and at their own expense, can inform the library, archive, and cultural heritage community about best practice in constructing online resources, and reaching relevant audiences in the process. (Terras, 2010)*

Terras states that online resources vary greatly, both in the quality of content and subject, and that the majority are of poor quality, many having been abandoned online. Terras does not define "poor quality", she only mentions that the best resources are accurate, authoritative, objective, current, of a good coverage and have unique information related to the collection.

Quality

The quality of an online resource depends on several factors. If we are judging a resource as a whole, we must use a scale of judgement which is different from that used when individual records are examined. For example, if a resource is the only information source that can be found, then it has a lot of importance. This affects the judgement over the resource, and the list of attributes for good resources by Terras can then be applied. But even if the resource as a whole is not that good, it might still have some value. An abandoned resource might also contain unique and valuable information within it. If the overall quality of a site is low, there might still be some parts with information that is not found anywhere else. The resource might have a subjective perspective that offers a new point of view to the topic. Thus a resource does not have to be current in order to be a good resource. It does not have to be objective nor have a good coverage. If an online resource is not trying to be a museum and does not follow museum standards, then it must not be judged according to those standards. A messy, ugly and abandoned site might have a unique perspective or an image that cannot be found anywhere else or a missing piece of information that makes it valuable as a online resource.

Current situation by examples

I have chosen arbitrary examples from real online materials that demonstrate some aspects of the current situation. Although arbitrary, these examples make it clear how technical implementations affect the visibility and re-usability of online materials, how images can work as linking elements between different sources, and how different are the attitudes that heritage organisations have towards

sharing their materials.

James Bond's toy gun

When Europeana is queried with the term "james bond", one of the first results is a photograph about James Bond's toy gun(Europeana, 2011). The image comes from the archives of the Nordiska Museet, and there is a quite typical collection data description written in Swedish attached to it (Nordiska Museet, 2011). This information is strictly limited to a technical description of the object, i.e. measurement and texts found in the item.

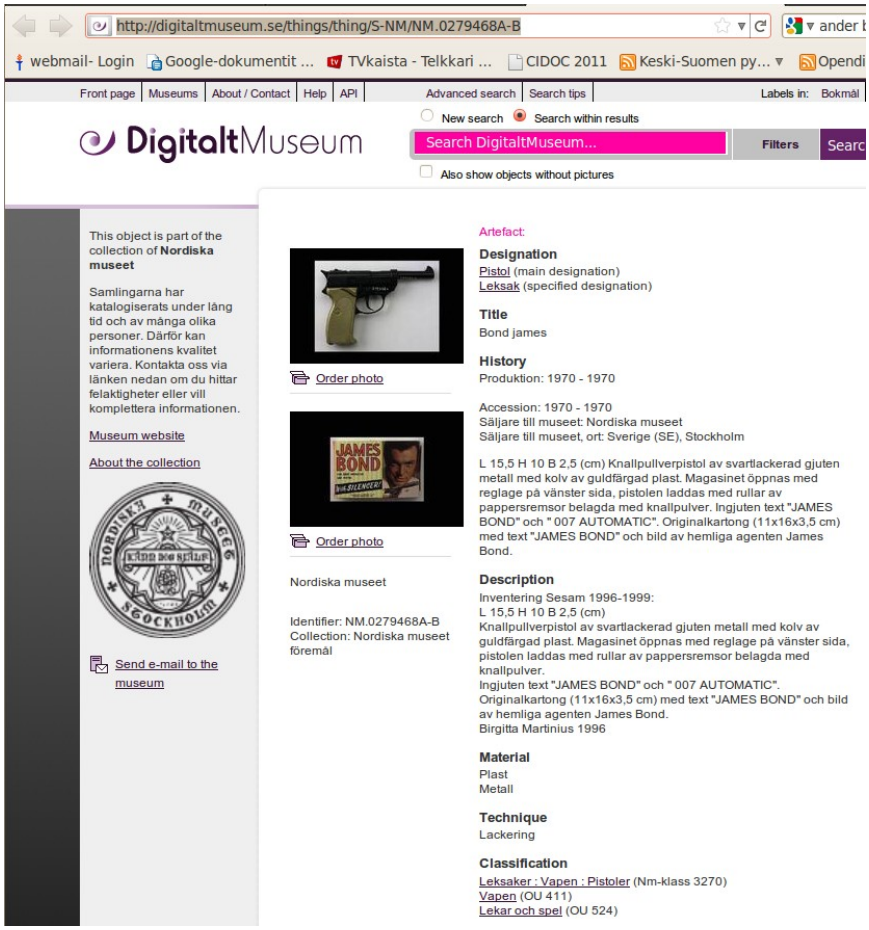

Illustration 1: The gun in the Nordiska Museet's online database.

Next, the same image and source were queried with the Google search using the following terms: "james bond pistol", "james bond toy gun" and "james bond leksak pistol" (Swedish). None of the searches produced an Europeana page or a Nordiska Museet page. English terms did not work because the description was written in Swedish, and the Swedish term for a toy gun (leksak pistol) did not work because the words "leksak pistol" were not present in the description text. Only query with a very specific term "james bond knallpulver pistol" – the term used in the description – was able to find the original page of the Nordiska Museet but not the page in the Europeana. If the name of the organisation (Nordiska Museet) was added to the search query, then also the terms "james bond pistol" and "james bond leksak pistol" brought up the Nordiska Museet site.

The original source offered very little information about the toy gun. For example, there was no

information about the manufacturer or the model of the toy gun. Therefore, additional searches had to be based on search terms “James Bond” and “toy gun”. With the Google’s image search I could search similar guns by comparing them to the images provided by the Nordiska Museet. A similar gun was found from a site called CollectToys.net. The site describes itself as follows: *CollectToys.Net is an online archive of images and information about vintage toys and collectibles from the 1950's through the 1980's.* The name of the manufacturer (Lone Star), a release year (1964) and an image of the gun was found from the site. More information about Lone Start could then be found from Wikipedia. Also, a virtual collection of cap guns, containing other toy guns made by Lone Star (http://www.nicholscapguns.com), was found. The most detailed images could be find from Ebay, where a gun set was for sale. But maybe the most interesting site - at least when it comes to contextualising the toy gun - was a certain kind of person's toy biography, which also included James Bond’s toy gun. This page places a toy gun in a personal narrative of childhood. [http://wesclark.com/am/toys.html].

Illustration 2: James Bond cap gun on Ebay

The following information was found from external sources:

- The gun had a silencer, which probably is missing from the Nordisk Museet’s gun (there is no mention about a silencer in the museum’s site).
- The gun model was introduced in 1964.
- The gun is not a replica of the actual gun used by James Bond.
- The manufacturer of the gun is Lone Star.
- There were also a larger set available with handcuffs, a gun holder and a badge.
- The Nordiska Museet seems to be the only museum having the gun in its collection.

The case of James Bond's toy gun reveals some interesting points. First, according to searches made, the Nordiska Museet is the only museum having the gun in its collection, but it is almost impossible to find it with Google, mostly because of the language barrier. Second, this case shows the importance of

sharing images in the cultural-historical field. The Nordiska Museet's information about the James Bond's toy gun was sparse and offered almost no contextual information. Additional information could be found only by comparing the image provided by the Nordiska Museet with images from external sources. What linked the museum resource to more informative sources was the image not the metadata provided by the museum. There are no semantic solutions for this kind of search problem. The only valid reference to the toy gun was the image since there was no model or even manufacturer mentioned.

National costumes of Finland

The best collection of national costumes of Finland is located in The Craft Museum of Finland. The museum is one of the first museums in Finland, having released their collection data online in 2006. However, there have been no resources to update the system since (Kotilainen, personal communication, 7 April 2011). The museum has several different collections online, all using the same technical platform.

The database shows no individual URLs for the user of different items – all the activities are under the same URL. Because of this, the museum's database disappears entirely within the deep web: its collection data is completely invisible to search engines. This means not only that none of the information in the database can be found with Google, but also that it is not possible to create links from external pages to any specific item in the database, leading to total isolation of the resource. Nothing can be found from outside, and nothing from outside can point to the database.

The Craft Museum has also another online resource dedicated to national costumes. The kansallispuvut.fi site was launched somewhere between 2004 and 2005, and it is now about to be renewed. Despite the age, the site is an excellent resource for anyone interested in Finnish national costumes. However, the site is implemented with frames (W3C, 1999), which are problematic for search engines since they do not correspond to the conceptual model of the web (Google, 2011). The problem with frames is that pages are indexed as individual pages, not as frames. The user might enter an orphan page that was not designed to be viewed without the frames around it. Site designers usually fix that by a small script that forces the browser to jump to the main page of the site. The unfortunate result of this is that the user now enters a page, which may be totally irrelevant to the original search query. This kind of frame structure with an “orphan page” script makes linking to subpages impossible since the user following the link always ends up with the main page of the site.

The “orphan script” solution is used also in the kansallispuvut.fi site, thus preventing linking to the subpages. Although pages are indexed and appear at the top of the search result list, there is no way to get from Google's search results to a subpage. For some reason – probably by mistake - the script is not included with all the pages. This leads to even more confusing result: some of the pages can be found and accessed with Google and some not. For example, the national costumes of Kurikka and Puulavesi both appear at the top of a search result list by Google, but the links behave totally differently. The link to the national costumes of Kurikka leads to the correct page in the kansallisspuvut.fi, but the link to the costumes of Puulavesi goes to the main page of the site without a mention about Puulavesi. This also means that it is possible to make a link to the page of the national costumes of Kurikka, but a link cannot be made to the page containing information about the national costumes of Puulavesi.

Since some of the images on the site are not owned by the museum, an attempt has been made, by employing a small script, to prevent their downloading (Hintsanen, personal communication, 11 April 2011). However, there is no way to prevent users copying what they see in their screens, which renders

this method unworkable. Modern browsers do not seem to respect the script, and it took less than 4 seconds to download all images on the site with a file retrieving program called Wget (Nikšić, n d).

This case demonstrates the situation where a technical platform hides the information from search engines or makes accessing information difficult. The choice of a technical framework for an organisation's online facilities determines how isolated or how open their online presence will be. For large museums, which possess the necessary ITC resources, technical solutions can be adjusted to the task by the museum's online policies, but for a small organisation there is a risk that instead of the museum itself creating the museum's online policy it is the technical platform that does it.

A cultural environmental portal

The National Board of Antiquities (NBA) of Finland has published a cultural environment portal, which holds information about Finnish relics, churches and other heritage sites (NBA, n d). The information consists of images, textual descriptions, protection decisions and geoinformation. The portal has explicit URLs for each record, thus allowing direct linking to the records. However, the content is not indexed by search engines since there is no site map or index page with links where the search robots could start their indexing. Nevertheless, there are some records that are indexed regardless of the non-search-engine-friendly system design. For example, the record of the church of Vuoksenniska is found with Google search because there is a link to it in the Wikipedia (Wikipedia) and also in a discussion forum related to church organs (Polso, 2009). Those links enabled Google to index the record page, and therefore both the record and the images on the record page can be found with a regular Google search.

This example from the National Bureau of Antiquities portal shows an interesting situation. The information of the database is online, but it remains isolated until someone publishes a link to it. This linking exposes the information to search robots, and thus makes the information public in the sense of accessibility through generic search engines. The other important part of this case is about URLs. The URL provided by the heritage portal is a technical URL. It is long and platform-specific, which means that changing the platform breaks all the links to the current database. This might happen without the organisation even realising that they have made their network of information inaccessible. This is another example of how design decisions affect - whether intentionally or not – the publicity and the usability of cultural heritage information.

Bowden Spacelander

The Brooklyn Museum has licensed their collection images under Creative Commons and provides an open application interface for fetching data to the user's own applications [https://creativecommons.org/tag/brooklyn-museum]. Among these images is the image of Bowden Spacelander, a rare bicycle with its futuristic design and fibreglass body. At the beginning of 2011, the Brooklyn Museum's page about Bowden Spacelander (Brooklyn Museum, a) was found on the fourth page of Google search results. Back then the bicycle was not on view in the museum, and there was very little contextual information in the online database. Currently the bicycle is on view, the page is updated and there is another page related to the exhibition showing the bike. This page makes it to the first page of the search results.

My first query, which revealed many information sources about the bike, was with the term "Bowden Spacelander". The National Bicycle History Archive of America by Leon Dixon (Dixon, n d) was the first hit. The site's web design does not follow the latest web design trends, but it has a scanned copy of an article from the Cyclist magazine concerning Bowden Spacelander. The site also describes how it is

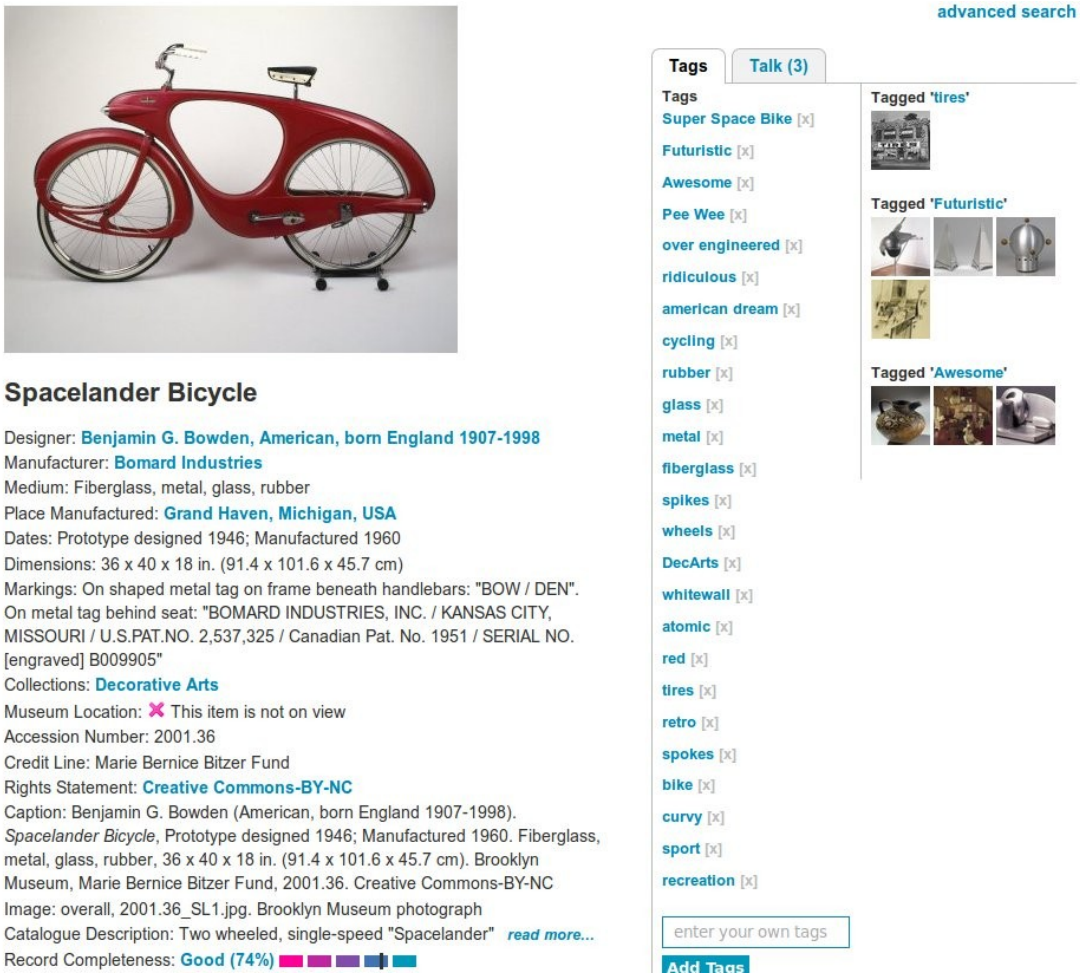

Illustration 3: Bowden Spacelander in the Brooklyn Museum's online database.

possible to tell the difference between the original bike and any replicas. Next I made an image-based search with an image found from the Brooklyn Museum's site. I used Tineye, which advertises itself as a reverse image search tool. Tineye found 9 matches, but a couple of them were images from the Brooklyn Museum itself. One of the matches was a blog called Mygreenweek, which has an article about Bowden Spacelander (Mygreenweek, n d), a little information about its designer and a lot of admiration for the design of the bike. The blog also shows one additional image, and there is a link to the National Bicycle History Archive of America. Although the information provided by the blog is not impressive, it still is relevant information. The main point with the blog is that it was not found by a textual search using the term “Bowden Spacelander”. The image from the Brooklyn Museum's collection was the link between these two resources.

Several observations can be made about the case of Bowden Spacelander. First, the most interesting and most informative material (magazine article, more detailed photos, blog texts, other similar bikes) were found outside the museum. Second, the image published by the Brooklyn museum was used on several pages, and those pages could be found with an image-based search. This is important since this works also in other direction: when the image from the blog is used as a query in Tineye, the resource of the Brooklyn Museum is found. And third, the image was not used the way intended by the Brooklyn Museum (embedding) but by copying it. And finally, none of the resources found that used the image

of the bike from the museum's collection provided a link back to the collection page. Again, only an image-based search could provide a link to the source of the image.

The Brooklyn Museum was not the best source of information in this case. However, it did contribute to the open information resources by publishing a photograph of the bike, which was then used on several sites. By using the image provided by the museum in an image-based search it was possible to find the resource (Mygreenweek blog) that did not appear with a textual search. This shows that it is possible to use images as a part of a linking structure between institutional and non-institutional resources. It is likely that image-based searches will be developed further. Museums could use image-based searches not only for tracking their images but for collecting new contexts in which the images have been used.

Măreţele maimuţe and William Holbrook Beard

Another example taken from the Brooklyn Museum's collection of images is the image of a painting called "For What Was I Created" by William Holbrook Beard (Brooklyn Museum, b). When this image was used as an input for Tineye, it found one similar image, namely, a cover of a Romanian edition of a book called Great Apes by Will Self. The Romanian title of the book is Măreţele maimuţe. The language of the page (Rogozanu, 2009) on which the image was located is Romanian and there was no mention about the cover art on the page. The only link between the Brooklyn Museum's resource and the book cover was the image of the painting by William Holbrook Beard. As in the previous case, the search also works the other way round. By feeding the image of the Romanian edition's cover art to Tineye, a picture of the Beard's artwork on the Brooklyn Museum's site is found.

Illustration 4: The cover of the Romanian edition.

One may ask whether this link between the cover art of the book and the actual painting is relevant. After all, the book is not about Beard's art, and it is not even an art book. Therefore, from the museum's point of view this link might be irrelevant. But if we consider the situation where a person finds the cover image of "Măreţele maimuţe" and gets curious about the cover's artwork, then the possibility of finding it by using the image as a search term is valuable.

Goya's paintings

There is no problem in finding images of Goya's works with search engines. Google's image search finds more than a hundred images of Goya's "Blind Man's Bluff" (La Callina Ciega in Spanish) alone. Many of those images are from commercial sites, such as ownapainting.com or allposters.com, offering posters or hand painted copies of Goya and other famous painters' works. Some of the images are from virtual collections like franciscodegoya.net

When a regular search is made with the term "Francisco Goya", the first in the result list is the Wikipedia, the second is a virtual collection of Goya works (franciscodegoya.net) and the third is the

Metropolitan Museum of Art's page of Goya, which is the only museum that appears in the first three result pages. Museo Nacional del Prado's online collection database has relatively good information in it, there are high resolution images available and there is even an audio tour. Even that is not enough to lift it to the top of the list when searching Goya's works. When searching his works "Blow", "The Third of May", "Nude Maja" and "Blind Man's Bluff", Prado is not listed in the first two result pages for any of the searches. The only exception is "The Third of May", with which Prado's site is listed on the second page.

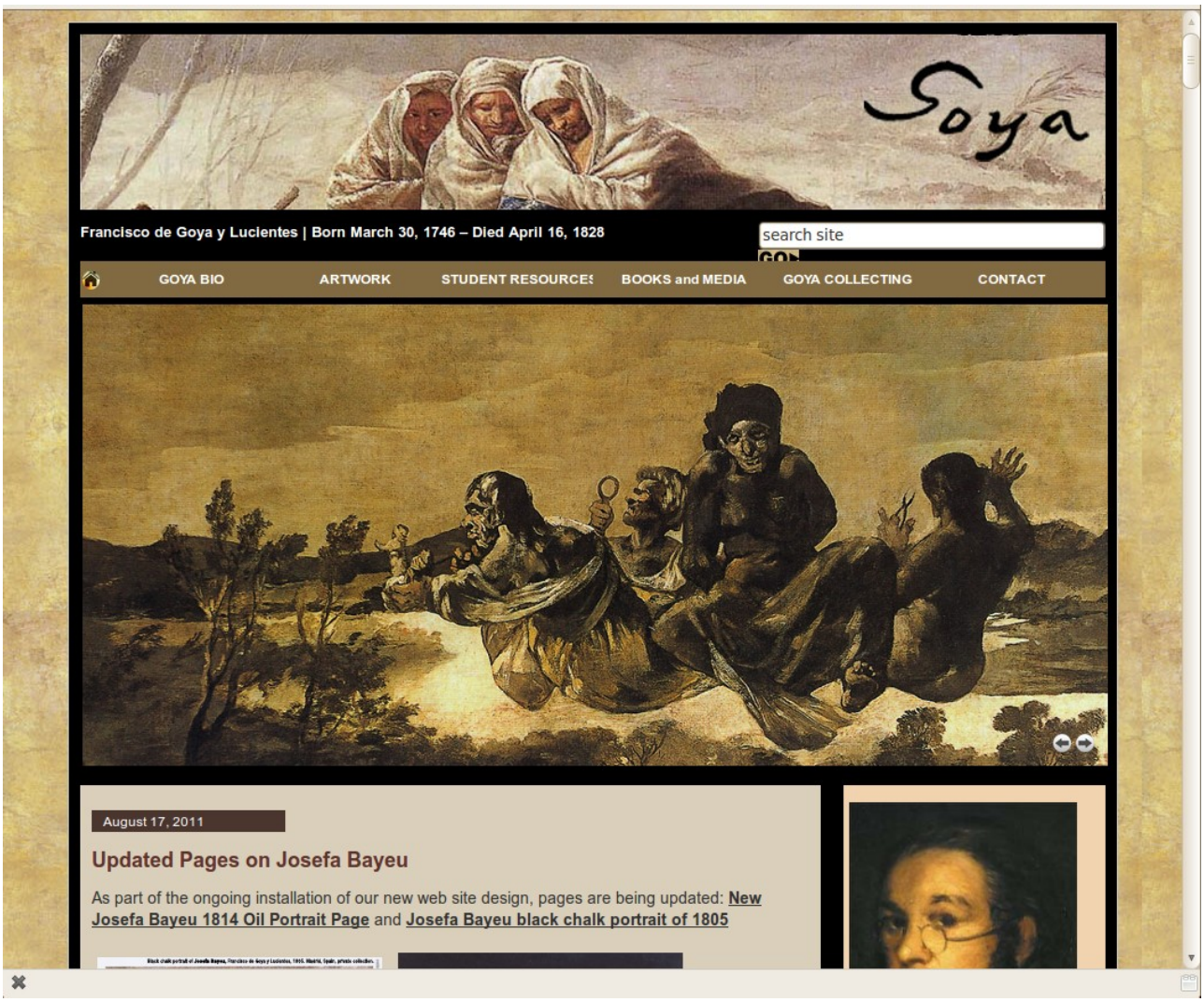

Illustration 5: The Goya site

Non-institutional sources dominate the search results of Goya's works. It is enjoyable to browse through that painter's works in franciscodegoya.net. There is no museum that could compete with this collection when it comes to the number of works. Another remarkable Goya resource is eeweems.com/goya. The overall feeling of the site is very "institutional" and professional. The site has sections like *bio*, *artwork*, *books,* and there is even a section for helping students. The books section has introductions to several books on Goya, there is a biography of Goya and the time line of his works; in the artwork section there are about 80 works with pictures; and in the links section there are more resources that are linked. In addition, there is an up-to-date news section about (real world) exhibitions and new articles; there is an article about Milos Forman's movie "Goya's Ghosts"; and there

are even instructions on how the site should be cited in academic works. The information on the site relies heavily on quotes from Goya literature. The quotes are correctly referenced so that the reader can find their original sources.

The case of Goya's paintings shows that in certain situations culture heritage organisations' online resources are almost completely overruled by non-institutional resources. There is no museum that would have all the paintings of Goya. The collection of all paintings by Goya can be created only virtually. This is the fundamental difference between non-institutional web resources and online resources of heritage organisations such as museums.

American Bison at the Castle

The image of two bisons in front of the Smithsonian Castle somewhere between 1887 and 1889 was a Smithsonian snapshot when Smithsonian news pages were accessed on 12th of August 2011 (Porter, n d). The image on the page comes from Smithsonian Institution Archives, but there is no link to the archive record. The image on the snapshot page is titled "American Bison at the Castle". However, when using the phrase as search terms in the Smithsonian Institution Research Information System or Institution's collection, the search did not find the image. Only when Tineye was used for an image-based search for the image, the same image could be found from a blog called "The Bigger Picture", which happens to be a blog of Smithsonian Archives. This image has a link to the actual record of the

Illustration 6: An image in the Smithsonian Snapshot without source.

archive in SIRIS (Smithsonian Institute, n d). When the same image was used as a search term in Google's image-based search, even more resources were found. This case shows that even heritage organisations themselves might lose the source of their own images. With image search, the original image could be tracked and found.

Discussion

All these examples (Isaksen, Heath, author) are hand picked, and one must be very careful when making generalisations based on these examples. But I believe that these examples demonstrate the current situation very well: when people search information about cultural heritage objects, it is done with the help of search engines, and heritage institutions perform often weakly in page ranks. But why is the current situation like this? And more importantly, is it good or bad?

Technical reasons are at least partially responsible for the weak presence of museums' online materials in the search engine universe. A major part of this problem is related to the concept of deep web. The deep web, or hidden web, means that part of the web content that is out of reach for generic search engines. Dynamic websites, password protected sites and non-textual content cannot be found by search robots and therefore they are not indexed. No one knows the size of the deep web, but it is estimated that the deep web is hundreds of times larger than the conventional web "on the surface". A typical deep web site is a museum’s online collection database. To avoid being trapped by the deep web, requires some expertise, which in small organisations with limited resources is rarely available.

The second reason for the weak search query presence is the lack of interesting information or material. If the only information that museum can provide about a collection item is measurements, technical description and the year of acquisition, then it is hard to compete with more contextualised resources. The third reason for not being at the top of the search results is related to advertising, or more precisely, lack of it. For example, commercial sites selling art posters advertise themselves actively. For these sites, high page rankings mean more business opportunities. Even if these sites have very limited information, they have enough links and traffic to lift their rankings above the rankings of heritage organisations.

Whether weak search engine ranking is a problem or not depends on several factors. Firstly, it must be acknowledged that in some situations museums perform very well. For example, when operating in a small language area, institutional materials appear high in the search results since very few sites compete with them. But when the page ranking of the museum is low, the question to be answered is: how is the museum able to profile itself as an authoritative information source if it is not found at all?

The importance of images

The cases presented here have shown the importance of images as linking elements between different kinds of sources. Images have many benefits over textual information. They are easy to share and use in different contexts; they are mostly language independent; and, lastly, they are wanted by the users. When there is very little information available, images provided by a heritage organisation might be the most valuable source for someone searching information. The context of an image is not lost: image-based search provides 2-way linking of materials without any manual intervention.

Images provided online by heritage organisations are mostly digital surrogates of collection items. Fiona Cameron criticises those who see digital surrogates merely as carriers of the message of the original object (Cameron, 2008). In a way, a digital surrogate is just a link to the original item without meaning of its own when it is used as a reference for a collection management purposes. But that

changes when surrogate is taken out of that context. When it is leaked outside of this framework, the link between the original and the replica is broken or at least the importance of the link is decreased. The surrogate now has its own life: there can be alterations, it can be used just as an illustration for something totally different or it can be combined with a new context. When a blogger uses the photograph of Bowden Spacelander provided by the Brooklyn Museum, the image is no longer pointing to the bicycle in the collection. The meaning of that image is now dependent on the context. At the minimum, the image is a image of a futuristic bicycle. At its most far-reaching form, the image refers to industrial development, the use of new materials (fibreglass), innovative design and to a story of a commercial failure. In other words, the image points to the concept of Bowden Spacelander.

Sharing or just displaying?

When a museum resource is finally found, the question remains whether the materials found can be re-used or not. Can I download the image of a toy that I used to have as a child? Can I post this same image to my blog and share my memories with the rest of the world? Should museums allow or even encourage re-use of their materials or should they try to prevent that? The question of sharing and re-using heritage materials is complex. There are legal, moral and financial issues involved (Eschenfelder & Caswell, 2010). I will discuss two fears in more detail: the fear of losing authority and the fear of re/de-contextualisation. Nina Simon has discussed the fear of losing authority in his blog article. She summarises this fear as follows:

> *One of the primary fears museum professionals (and all professionals) have about entering new relationships with audiences is the fear of losing control. For hundreds of years, we've owned the content and the message. While we may grudgingly acknowledge the fact that visitors create their own versions of the message around subsets of the content, we don't consciously empower visitors to redistribute their own substandard, non-authoritative messages.*

She points out that one must make a distinction between control and expertise: having control does not mean that there is any kind of expertise present, and having expertise does not require any kind of control.

A heritage organisation has to realise that when they expose images or information to the search engine universe, the control of the material is practically lost. Heritage organisations do not have control over the algorithms search engines use, they don't have a say on what kinds of images (see illustration 4) or links are shown parallel in the search results, and if materials are interesting, they will be re-used no

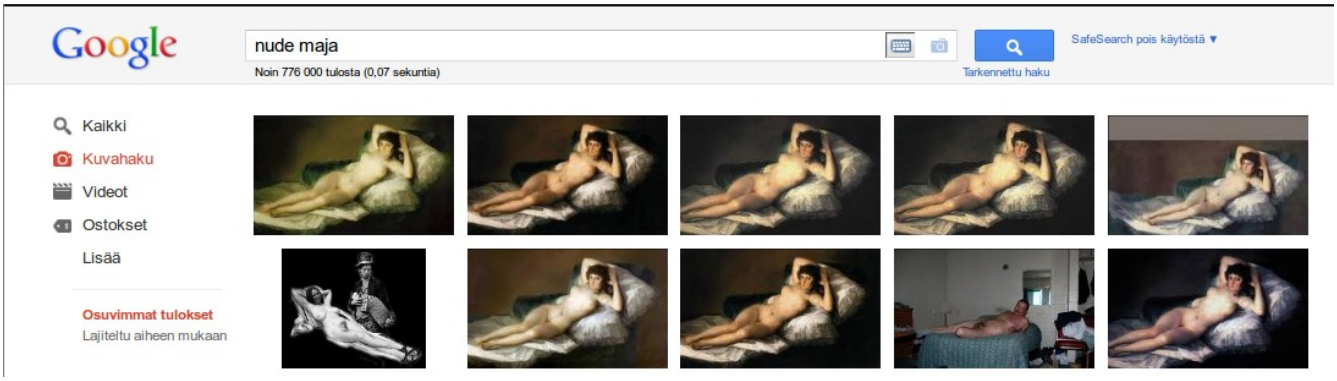

Illustration 7: Goya's Nude Maja in the Google's image search with SafeSearch off.

matter what the copyright status or restrictions are.

The other common fear is the fear of re/de-contextualisation of the materials. Museums are afraid that if they allow public use of their images, they might be used, for example, in an inappropriate context or

in a way that hides the original context. The key to the problem of re/de-contextualisation is in the way the museum defines its responsibility regarding the published material. The museum may see itself ultimately responsible for all possible uses of the materials, whether authorised or unauthorised, now or in the future. In a global, digital network this kind of responsibility is almost impossible to enforce. Global access, with the easiness of digital copying and digital altering renders this approach unworkable. A narrower viewpoint holds that the museum is responsible for their own usage of the material only. What happens “outside” is out of the control of the museum and therefore the museum cannot be held responsible for that.

Museums' policies can be divided into three categories, according to their attitudes regarding re-use: i.e. Virtual Display Case, Cultural Property/Regulated Access, and Cultural Remix (Eschenfelder & Caswell, 2010). The virtual display case works like a virtual glass display. It shows the item but prevents any copying and altering of it. Due to very flexible copyright practises among internet users, this approach has serious difficulties. The only way the approach can be made feasible is by practically destroying the image with watermarks. This prevents its re-use, but it also prevents one from viewing the image for one's enjoyment.

A discussion has taken place in a Finnish history forum called Agricola about watermarking policies of some Finnish heritage organisations. Kuvakokoelmat.fi by National Board of Antiquities and arjenhistoria.fi by the consortium of several Finnish museums have adopted the Virtual Display Case by watermarking heavily their images. There are supportes and opponents in the discussion (including the author) but the main tone of the discussion is very critical towards watermarking policy. One opponent even speaks about cultural vandalism and equates watermarks with graffitis in the walls of heritage targets (Onnela, 2011). A supporter states that a collection is property of a museum and the museum has no obligation to publish it online freely. A response for this was that if the museum is publicly funded, there is at least a moral obligation to do it.

Illustration 8: Watermarked example images from kuvakokoelmat.fi (left) and from arjenhistoria.fi (right).

From the museum's point of view the virtual display case model might look like an ideal solution. Materials are online but the control remains in the museum. However, the message send by the

museum is the message of the owner. The collection is property of the museum and the role of an online visitor is to visit, not to participate. The question to be answered is: Is a museum entitled to practically destroy a digital surrogate in order to prevent its use?

Flickr commons, Smithsonian Commons, Brooklyn Museum and The Yale Digital Commons are real world examples most close to the cultural remix model. Materials are classified as “public domain” or “no known copyright restrictions”. Material is provided for reuse similarly accepting the fact that the control of the material is mainly lost, at least when it comes to non-commercial use. The “no known copyright restrictions” moves part of the responsibility to users. This way images can be shared even when there is no permission from copyright holder that is not known.

Solutions

What solutions are there? Firstly, there are technical solutions for technical problems. When a museum exports its collection database to the net, it should provide a site map that would help search engines to index the site. The museum's online database should use non-technical and permanent URLs, ensuring that links from outside remain functional despite platform changes. Secondly, there are non-technical solutions. These solutions are based on existing practises of Internet and re-evaluation of the role of the museums.

In museum literature and museums' informatics literature the basic assumption is that it is the museum that has an active role. The museum sets the framework where everything happens. People can participate, produce content and tag items but mostly in terms set by the museum. From museum-centric perspective the solution would be something that museum would do: semantic mapping of collection data, Twitter and Facebook accounts for the museum in order to have more links, a blog, or a bunch of fanny games in the museum's website. But if the museum is put aside for a while, then we have a different kind of perspective; A museum could also be a passive content provider, allowing people themselves to invent what to do with the museum’s materials. Passive role means also that fewer resources are used. This does not mean that heritage organisations should stop making crowdsourcing projects or stop creating educational online environments. The role of a passive information/material provider and that of an active/authoritative heritage organisation do not exclude each other. Instead, they serve people in different situations and with different goals.

In this model, the museum sees itself as a resource that admits that there are other - and sometimes much better - sources of information outside the museum. Instead of encapsulating the museum’s online content, the content is shared so that others can combine it with different resources and enrich the content. It is like user-generated content creation or crowdsourcing outside of the museum context. The task is to bring digital heritage to everyday life. This does not necessarily provide better search engine visibility for the museum itself but it gives deserved visibility to the materials of the museum. What is needed is an external knowledge management model. A museum can not control the external information but it can decide – and manage – its own role in the network of information.

Conclusion

Museums see themselves - and they are seen - as trusted sources of information. At the same time, museums do acknowledge that their information about their own collection items is many times non-contextualised and sparse. If museums have problems in contextualising their collections and if that context already exists outside the museums and if the connection between these could be made automatically, then at least this is something that museums should study.

Instead of squeezing information to strict categories, heritage institutions might try the wild way by adopting the existing Internet culture for sharing. Copying and using one's materials can be seen as a positive sign. It can mean that an organisation’s materials are interesting. And in the other way round: if no-one is copying one's materials, then those materials are not interesting. Museums should encourage people to do some cherry picking. By selecting something that interests them and placing that to their own blog or website makes it more visible and more contextualised. There are also enthusiasts that are able to do much more. They can collect information and materials from several sources and combine them as a new resource that surpasses the original resources both in quality and coverage.

Is Google really the largest museum in the world? Google is a gigantic voting, re-contextualisation and marketing machine that connects information and materials without caring about academic degrees, institutional status or proper context. Google is a Wikipedia done in a large scale, with unlimited perspectives and without any publishing policy. Museums are not in the position to define the rules for this particular kind of museum. Only thing they can define is whether they will be part of it or not. The revolution of heritage authority has already taken place. There is a large scale crowdsourcing going on, people creating, commenting and manipulating digital heritage. Museums may give their support to it or they may ignore it but they cannot prevent it from happening.

Brooklyn Museum. (a). Bowden Spacelander. *Brooklyn Museum's database*. Retrieved from http://www.brooklynmuseum.org/opencollection/objects/2587/Spacelander_Bicycle

Brooklyn Museum. (b). For What Was I Created? *Brooklyn Museum's database*. Retrieved from http://www.brooklynmuseum.org/opencollection/objects/984/For_What_Was_I_Created

Cameron, F. (2008). Politics of Heritage Authorship: The case of digital heritage collections. *New Heritage. New Media and Cultural Heritage* (pp. 170-184).

Chiles, A. (2008). Lecturer bans students from using Google and Wikipedia. *The Argus*. Retrieved August 10, 2011, from http://www.theargus.co.uk/news/1961862.lecturer_bans_students_from_using_google_and_wikipedia/

Dixon, L. (n.d.). THE 1960 BOWDEN SPACELANDER! *National Bicycle History Archive of America*. Retrieved from http://nbhaa.com/indexBowden.html

Eschenfelder, R., & Caswell, M. (2010). Digital Cultural Collections in an Age of Reuse and Remixes. *First Monday*. Retrieved August 11, 2011, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3060/2640

Europeana. (2011). Europeana search. Retrieved August 2011, from http://www.europeana.eu/portal/search.html?query=james+bond

Google. (2011). Webmaster Tools help Frames. Retrieved August 12, 2011, from http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=34445

Harley, D. (2007). Why Study Users? An environmental scan of use and users of digital resources in humanities and social sciences undergraduate education. *First Monday*. Retrieved August 11, 2011, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1423/1341

Heath, S. (2010). Diversity and Reuse of Digital Resources for Ancient Mediterranean Material Culture. In G. Bodard & S. Mahony (Eds.), *Digital Research in the Study of Classical Antiquity* (pp. 35-52). Farnham, UK: Ashgate. Retrieved from http://sebastianheath.com/publications

Isaksen, L. (2009). Pandora ' s Box : The Future of Cultural Heritage on the World Wide Web. Retrieved August 10, 2011, from http://leifuss.files.wordpress.com/2009/04/pandorasboxrev1.pdf

Marty, P. F. (2008). Changing Needs and Expectations. *Museum Informatics People, Information and Technology in Museums* (pp. 182-185).

Mygreenweek. (n.d.). CRAZY WEDNESDAY: WONDERFUL NEW ELECTRIC BIKE!!! *Mygreenweek*. Retrieved from http://senseup.wordpress.com/?s=bowden

NBA. (n.d.). Cultural Environmental Portal. Retrieved August 2011, from http://kulttuuriymparisto.nba.fi/netsovellus/rekisteriportaali/portti/default.aspx

Nikšić, H. (n.d.). Wget. Retrieved from http://www.gnu.org/s/wget/

Nordiska Museet. (2011). James Bond's toy gun. Retrieved August 2011, from http://digitaltmuseum.se/things/thing/S-NM/NM.0279468A-B

OCLC. (2010). *Perceptions of Libraries, 2010: Context and Community*. *Context*. Retrieved from http://www.oclc.org/reports/2010perceptions.htm

Onnela, T. (2011). Digitoidut kuvakokoelmat verkossa, kenen käytössä ja millä ehdoilla? *Agricola discussion forum*. Retrieved August 2011, from http://agricola.utu.fi/keskustelu/viewtopic.php?f=10&t=3460&p=17572#p17572

Polso. (2009). Alvar Aallon urkufasadit. *Urkufoorumi*. Retrieved June 2011, from http://p2.foorumi.info/urkufoorumi/viewtopic.php?t=163

Porter, J. (n.d.). American Bison at the Castle, 1887 – 1889. *Smithsonian Newsdesk*. Retrieved August 2011, from napshot/american-bison-castle-1887-1889

Rieger, O. Y. (2009). Search engine use behavior of students and faculty: User perceptions and implications for future research. *First Monday*. Retrieved August 10, 2011, from http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2716/2385

Rogozanu, C. (2009). Relax, mai sînt şi cărţi mişto pe lumea asta. Ideile şi evoluţionismul. *Voxpublica*. Retrieved August 2011, from http://voxpublica.realitatea.net/popcult/relax-mai-sint-si-carti-misto-pe-lumea-asta-ideile-si-evolutionismul-20773.html

Smithsonian Institute. (n.d.). Buffalo Behind Smithsonian Institution Building. *Smithsonian Institution Research Information System*. Retrieved August 2011, from http://siris-sihistory.si.edu/ipac20/ipac.jsp?&profile=all&source=~!sichronology&uri=full=3100001~!9175~!0#focus

Terras, M. (2010). Digital curiosities: resource creation via amateur digitization. *Literary and Linguistic Computing*, *25*(4), 425-438. doi:10.1093/llc/fqq019

W3C. (1999). Frames in HTML documents. Retrieved August 2011, from http://www.w3.org/TR/html4/present/frames.html

Wikipedia. (n.d.). Vuoksenniskan kirkko. Retrieved from http://fi.wikipedia.org/wiki/Kolmen_Ristin_kirkko