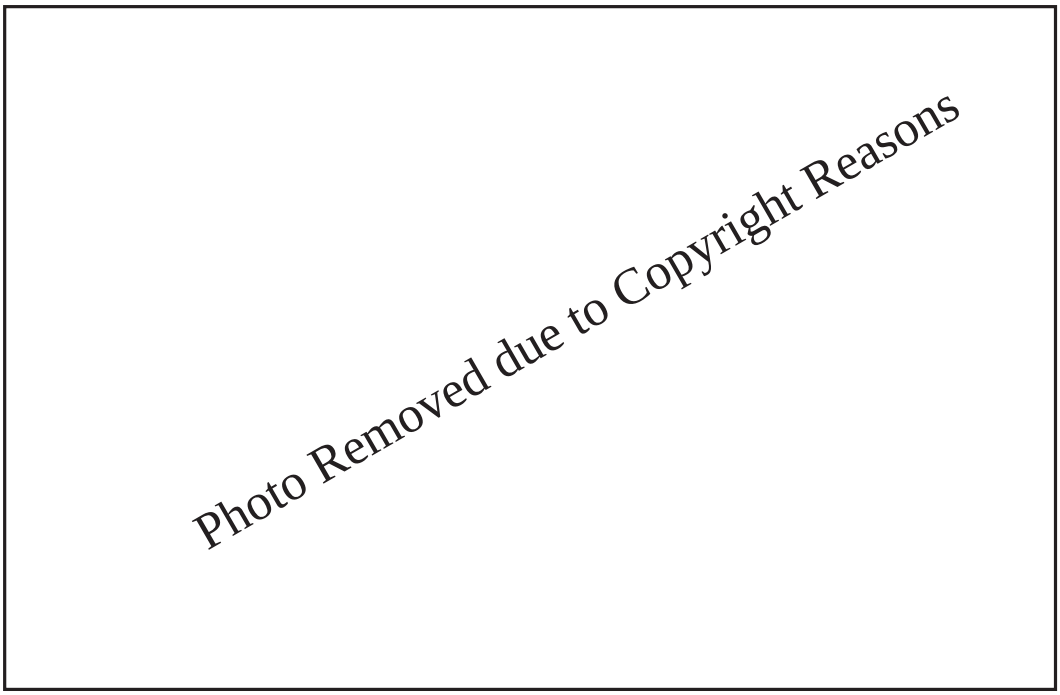
JYVÄSKYLÄ STUDIES IN HUMANITIES
187

Ari Häyrinen

Open Sourcing Digital Heritage

Digital Surrogates, Museums and Knowledge Management in the Age of Open Networks



JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN HUMANITIES 187

Ari Häyrinen

Open Sourcing Digital Heritage

Digital Surrogates, Museums and Knowledge Management in the Age of Open Networks

Esitetään Jyväskylän yliopiston humanistisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Historica-rakennuksen salissa H320
lokakuun 27. päivänä 2012 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Humanities of the University of Jyväskylä,
in building Historica, Auditorium H320, on October 27, 2012 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Open Sourcing Digital Heritage

Digital Surrogates, Museums and Knowledge Management in the Age of Open Networks

JYVÄSKYLÄ STUDIES IN HUMANITIES 187

Ari Häyrinen

Open Sourcing Digital Heritage

Digital Surrogates, Museums and Knowledge Management in the Age of Open Networks

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2012

Editors
Raine Koskimaa
Department of Arts and Culture Studies, University of Jyväskylä
Pekka Olsbo
Publishing Unit, University Library of Jyväskylä

Cover picture by Ari Häyrinen

URN:ISBN:978-951-39-4908-2
ISBN 978-951-39-4908-2 (PDF), ISSN 1459-4331 (PDF)

ISBN 978-951-39-4907-5 (nid.)
ISSN 1459-4323

Jyväskylä University Printing House, Jyväskylä 2012

ABSTRACT

Häyrinen, Ari
Open sourcing Digital Heritage. Digital surrogates, museums and knowledge management in the age of open networks.
Jyväskylä: University of Jyväskylä, 2012, 134 p.
(Jyväskylä Studies in Humanities
ISSN 1459-4323; 187)
ISBN 978-951-39-4907-5 (nid.)
ISBN 978-951-39-4908-2 (PDF)
Finnish summary
Diss.

I study the emergence of traditional online activities of the museums and their fears in the vivid and multifaceted Internet and how the open Internet is setting out to compete with museums. The second main topic of this work is the role of technology in the openness of digital heritage. Technology is always present when dealing with digital heritage, and understanding the technical framework is essential when evaluating, for example, long-term preservation issues or accessibility. I used three different kinds of sources in my study: literary, case studies and application development. First, I have read studies that are related to digital heritage: museology literature, museum informatics literature and information technology studies. Secondly, I studied online digital heritage materials. How can these be found and who are creating these resources? By whose terms can they be used? Why some materials are open and why some are not? Lastly, I have resorted to my own experience as a developer of a cultural heritage information system and as an open source enthusiast. I created technical proposals that demonstrate some solutions for documentation, openness and long-term preservation of digital heritage. I was able to spot several structural issues that negatively affect the ability of the heritage organisations to participate in the creation of open digital heritage. These disadvantages can be dubbed an institutional burden of official heritage institutions. The institutional burden is related to the practices of organisations, to the official status of these organisations and to the technology that they use when creating digital materials. I was able to demonstrate that technical implementations affect the visibility of the museum materials to search engines, the flexibility of the documentation and long-term preservation. The issues of the institutional burden are partly linked to these technical questions. A well-designed information system supports multiple documentation types, and openness of data does not require extra work, which, in turn, saves the resources of the organisation. Nevertheless, this requires changes in the information system design for heritage organisations.

Keywords: digital heritage, open source, open data

Author

Ari Häyrinen
Department of Arts and Culture Studies
University of Jyväskylä
Finland

Supervisor

Professor, Ph.D., Raine Koskimaa
Department of Arts and Culture Studies
University of Jyväskylä
Finland

Reviewers

Professor, Ph.D., Jaakko Suominen
Degree Program in Cultural Production and Landscape Studies
University of Turku
Finland

Associate Professor. Ph.D., Christian-Emil Ore
Department of Linguistics and Scandinavian Studies
University of Oslo
Norway

Opponent

Associate Professor. Ph.D., Christian-Emil Ore
Department of Linguistics and Scandinavian Studies
University of Oslo
Norway

LIST OF FIGURES

CONTENTS

ACKNOWLEDGEMENTS

My second choice of title for my work was “Liberating Digital Cultural Heritage”. Although I liked the idea of presenting myself as a liberator, I decided not to use that title. There were two reasons for that. Firstly, liberating is a strong word - especially when used in an academic study - because it contains a claim that someone or something is being held by force. And that is not the case with digital cultural heritage. There is no hostage situation there, merely a mix of misunderstandings, unwillingness and fear towards opening digital access to cultural heritage. Secondly, the field of cultural heritage is so wide and there are so many organisations and so many persons and agendas involved that it is beyond the powers of one individual researcher to liberate anything.

I would like to thank my supervisor Professor Raine Koskimaa for his understanding and guidance during my - sometimes curvy - path to the final work. Professor Heikka Hanka has offered me several opportunities to deepen my understanding about cultural heritage and ICT issues by inviting me to several meetings and occasions. Dr. Martin Doerr gave me some insight concerning ontological modelling and I thank him for this valuable information. Discussions during CIDOC conferences with Mika Nyman were fruitful source of ideas. In addition, the insightful comments of both reviwers of this thesis, Professor Jaakko Suominen and Associate Professor Christian-Emil Ore, were very helpful when I worked with the final version of this text.

I would also like to thank the Cultural Historical Section of the Jyväskylä University Museum and their personnel, especially Intendant Pirjo Vuorinen, for their time and input. Without their support this work would not have been possible. Fellow students and members of Taiku in the former Juomatehdas have created an inspirational and flexible environment which I enjoyed very much. Special thanks to Tellervo Helin and Juha Teppo for their support in the early stages of my transformation to an digitally orientated art historian. In addition, the 3D-Bridge was an important step for my dissertation and I like to thank all persons that were involved this project.

My research project was supported by the Finnish Cultural Foundation, the Kone Foundation and the University of Jyväskylä. In addition, the Alfred Kordelin Foundation sponsored my journey to the CIDOC 2011 conference in Romania.

Lastly, I want to thank my family; my mother, my sister, dear Sari and very dear (and very small) Aarni.

1 INTRODUCTION

1.1 Research questions and materials

> "Time and accident are committing daily havoc on the originals (of valuable historic and state papers) deposited in our public offices. The late war has undone the work of centuries in this business. The lost cannot be recovered; but let us save what remains; not by vaults and locks which fence them in from the public eye and use in consigning them to the waste of time, but by such multiplication of copies as shall place them beyond the reach of accident." –Thomas Jefferson, February 18, 1791

We are surrounded by physical objects, and we think that some of those objects have characteristics that make them so valuable that they should be preserved for future generations. That part of physical artefacts forms our tangible cultural heritage. Tangible cultural heritage is about fragile, often unique items that need constant caring by trained professionals and protection from the wearing effect of the hands of culture tourists. Moisture, dryness, heat, cold, dust, thefts and acid rain are constantly threatening the existence of these treasures. Huge amounts of resources are used for preservation and conservation of buildings, art works and archaeological findings so that they could be saved for future generations.

Currently, an increasing number of these tangible heritage objects have digital representations as images, videos or 3D-models, for example. In addition, tangible heritage objects are subject to digital documentation practises, which create connections between different items, information sources and concepts. Heritage institutions - museums, libraries and archives - are the main sources of digitised heritage objects. This new heritage is essentially global when put online. It does not wear out when used or when copied. Still, it is fragile and needs active preservation activities.

The Internet, as we know it today, is a frightening, fascinating, uncontrolled and even anarchistic network. On the other hand, we have heritage institutions with their strictly controlled collections and long traditions. When these institutions go online, there is an interesting collision of two very different kinds of cultures. This collision is manifested in heritage institutions as various kinds of

fears, like the fear of misuse of heritage materials, the fear of decontextualisation or the fear of loosing authority as an expert organisation.

Although institutional heritage organisations have a major role also in the field of digital heritage, they are not the only creators of digital heritage. 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 can compete and even surpass the institutional sources in search results, coverage and relevance. All these sources are mixed and connected quite unexpected ways when search engines organise the web according to our search criteria. This is the field where institutional digital heritage and non-institutional digital heritage meet and collide.

My research questions are related to this borderline between institutional and non-institutional digital heritage. I examine digital heritage from the outside of the heritage organisations. I use the term "everyday Internet" for that part of the Internet that people use in their everyday life, and search engines are an important part of that. I study the emergence of traditional online activities of the museums and their fears in the vivid and multifaceted Internet and how the open Internet is setting out to compete with museums. I see the digital surrogates of tangible heritage artefacts as an access and reinterpretation method. Digital surrogates are digital anchor points that can serve as references for participation and discussion. The availability of these surrogates is crucial for people to participate in digital heritage. The main question now is: Do heritage organisations want to be part of the open digital heritage or do they want to keep aloof in order to maintain their authority and control over their materials. What will happen once they have gone online and on whose terms this takes place?

From the set of heritage organisations - museums, libraries and archives - , I concentrate merely on museums, their practises and their role in what I call "open digital heritage". I have chosen museums as my main target for a couple of reasons, of which the most important reason is the difference in the way these organisations create and handle their collections. Library objects are usually able to "explain" themselves. There is no need take a photograph of a book and write a description about what a book is, for example. Most of the objects in the library collections have a title, a known author, and a clear topic, which makes cataloguing and information retrieval relatively simple. This also applies to archive materials, at least to a certain extent. The unambiguous cataloguing practices of these organisations reflect potential openness of these materials. In fact, many libraries have opened their catalogues as Open Data sets, and archives have their own initiative for open access called Open Archive Initiative.

However, museum collections can include almost anything from tractors and letters to art installations and toy guns. Being collected entities, there is something for everyone in them, which makes them interesting for general public. Still, a museum item needs an explanation or interpretation in order to be a cultural object. There can be several different explanations and interpretations depending on who is the contributor. It is also possible that these views are totally missing from the official collection data because of the lack of resources.

Moreover, the great variety of collections sets special requirements for museums' information management systems and ways that information can be made public. Together these reasons render museums as one of the main topics of this work.

I have excluded the questions of digitisation process itself like "what should be digitised?" or "what are the technical requirements for good digitisation?". Instead, I am merely interested in knowing when digital surrogates are made available online or why surrogates are not accessible at all. My argument is that heritage organisations - and especially museums - are in a new position when it comes to digital heritage and open Internet. These organisations form part of open information networks through search engines, willingly or against their will. The change from the "access Internet" to a "participatory Internet" has lead to a situation where heritage organisations have lost their privileges and become regular information sources competing against any other resources available in the search engine universe. They are confronting the challenges of the open Internet, where everyone can create information resources for search engines to find. The internal and the external are mixed, and more participants are taking part. Knowledge management, which was previously an internal affair, has a different meaning in the open, networked reality.

The second main topic of this work is the role of technology in the openness of digital heritage. Technology is always present when dealing with digital heritage, and understanding the technical framework is essential when evaluating, for example, long-term preservation issues or accessibility. Museums have long traditions in the management of collections and have their information systems in the form of collection management applications. These applications are designed for the internal work of museums, and they are also guiding museums on how to model and handle their data. There are solutions that are more open and more accessible than others, but determining that in practice can be quite complex. The other side of digital heritage technology is whether it affects documentation practices positively or negatively. Documentation gives the context a target - or at least this is true in an ideal case -, and it helps to locate the resource. Therefore, documentation practices and techniques, like semantic solutions, are an important part of my work.

I have used three different kinds of sources in my study: literary, case studies and application development. First, I have read studies that are related to digital heritage: museology literature, museum informatics literature and information technology studies. I have studied museum literature in relation to how museums see themselves, to their relationship to public and to their attitude to sharing. Museum informatics literature gave me insight of information system design practices and issues in the museum field. Information technology literature revealed issues and practices about computer-based information handling in general. Secondly, I have studied online digital heritage materials. How can these be found and who are creating these resources? By whose terms can they be used? Why some materials are open and why some are not? Lastly, I have resorted to my own experience as a developer of a cultural heritage information system and as an open source enthusiast. What makes building open informa-

tion systems for cultural heritage difficult in practice? Where do these limitations come from? From these sources I have tried to locate roadblocks and solutions for open digital heritage.

My work could be categorised under several disciplines. Currently, this work is published under digital culture, but it could have been published - at least partly - under digital humanities, human computing, museum studies, museum informatics or ICT. This interdisciplinarity is necessary in order to address the complexity of the open digital heritage. Traditions of museum work, practices of documentation, technology of documentation platforms, design principles of heritage ICT, ontological issues of semantic computing and long-term preservation issues of digital content are all aspects of digital heritage that must be covered in order to study the phenomena.

Although I primarily study digital heritage in the museum context, there are some aspects that can be expanded beyond this context. Archives concentrate on archiving, conservation and preservation of certain kinds of documents but are not necessarily physically exhibited. But if we suppose that the documents of an archive are made publicly available through the Internet, then most of what have been said here about museums and digital heritage also applies to the online materials of that archive. However, the difference between archives and museums is not always clear. Museums also have document archives, and both archives and museums have image collections, for example.

I have tried to write the main part of this work in a way that would make it possible to understand it without the need to read the Included Papers section. First, I reflect on the process of my work and how it was developed. Then I introduce several aspects that are related to the concept of digital heritage. Digital surrogates, search engine universe and open data movement are among them. From this introduction, I move to the problems of open digital heritage. These problems concern the creation, preservation and openness of digital heritage. In the Cases section, I present several online cases that exemplify several situations of digital heritage, such as the role of technical implementations in search engine visibility, the role of freely usable images and non-institutional information sources that are competing with the official ones. Lastly, I present, with demonstrations, some views about the design principles for more open digital heritage information systems.

1.2 Evolution and methodology of this work

I have been involved with open source software several years as a user and later as a developer. At this time I have internalised - or at least become accustomed to - openness in all levels. I can participate in open source projects, and I can modify, use and distribute open source applications. Therefore I was surprised of the restrictions set by some heritage organisations for their online materials. There are significant exceptions to this exclusivity of online heritage materials, but, to

a person with my background, the online presence of museums had created environments that on viewing seem almost hostile when it comes to accessing and re-using heritage materials. I wanted to find out reasons for that and, further, I wanted to study whether there would be a niche for open digital heritage in the sense of open source movement.

My work started during an EU-funded project called 3D-Bridge. It was a one year project and the aim of our sub-project was defined as following:

> The aim of our subproject was to develop a new type of documentation method for cultural-historical reconstruction. By combining database approach with hyperlinking and virtual reality application, it was possible to develop a tool that supports group work and is flexible enough for storing different types of documentation. The area of Seminaarinmäki was used as a test case for the documentation model. We gathered new information about the area by modelling its buildings and storing metadata to the database. (Lonkila, 2006)

We wanted to develop open methods for 3D-documentation for cultural historical use. During the project a method for modelling architecture with open source software was developed (Tammisto and Häyrinen, 2006). The project revealed some difficulties in the field of cultural historical documentation. Although 3D models can be technically combined with a documentation in a database, the actual problem was how to produce and organise the documentation itself. Current solutions were item-centric and rather rigid. We needed solution that would not be limited to items and that could be used to document very different kind of targets like images, buildings and campus areas. The result was a software prototype called IDA3D. Developing the software revealed how complex issues there were in the cultural historical documentation. Especially data modelling caused a lot of problems and most of the problems in this area remained unsolved during the project.

After the 3D-Bridge I started individual work for a doctoral thesis. The 3D-Bridge project clearly demonstrated that more flexible, and more open way for documenting cultural historical targets was needed. My first approach was to develop a software framework that would provide a flexible and semantically meaningful documentation model. In addition, the framework should provide an open application programming interface that would allow third parties to use information.

However, during the process the focus of my practical work changed from software development to a study of open heritage information systems and further to a study of principles of open heritage. Reasons for this change were both theoretical and practical. One of the continuous problems with cultural historical documentation is the lack of resources. I realised that yet another software project would be hardly able to solve that problem. What was needed was a deeper analysis of what are the components of digital heritage and how these components affect to the openness of the digital heritage.

The rest of the reasons for giving up the software development as a main approach were more practical. If an application has users, then there must be

some kind of support services and I failed to see myself as a support service provider. Commercial support services exists for open source software but that requires a user base that is able to pay for support, which I saw unrealistic. After all, the lack of resources - both human and financial - was the initial problem. A second practical problem was the importance of user interface. A user interface is the contact point where a developer and an user meet. While I was developing the first version of the software, I found myself using more and more time on user interface issues. Issues that were not part of my work. Creating a good and usable interface is a time consuming task. All time spent on user interface issues was away from the actual development of the core application.

Because of the reasons mentioned above, I decided to change the course. I continued software developing but instead of aiming to a finalised application, I started to search generic developing principles for open digital heritage tools. I used programming in order to demonstrate some principles that were essential for long-term preservation, openness and flexibility of digital heritage information systems. In order to define good design principles for open heritage applications I had to resolve of what kind of framework museums were operating when they went online. I was not interested of what museums might do in the future or what kind of online strategies were written but I wanted to see the current, real world situation. As generic search engines are main entry points to the Internet, I focused on what was found with Google instead of using museum's own search engines. In many cases the results were surprising. Although the organisations themselves were easily found with Google, this was not the case with their online materials. In some cases the whole official digital heritage sector was non existent when searching information about heritage targets. Instead, other sources of information were found that provided information needed. There were cases were museum had deliberately prevented search engines for indexing the site. In the world where companies, associations and individuals are using all kind of methods to improve their search engine ranking this kind of activity is curious. That raised the question about heritage organisations search engine visibility and museums' role in the world of open digital heritage in general.

The role of unofficial sector was surprisingly strong when compared to the online presence of official digital heritage. I realised that digital heritage is much more than digital surrogates produced by heritage organisations. By studying museum and museum informatics literature I was able to explore reasons for this digital muteness of museums. In addition, I analysed some online cases. I selected them by their exemplariness. I have selected examples that are controversial (Tipitii exhibition), that raises strong feelings (Anne Frank museum), that demonstrates how technical implementations affected the accessibility of the data (Finnish national suites), that show how museums are bypassed as an online information sources (Goya) or that exemplifies the structural disadvantages that museums have (C64.com). I studied these online resources qualitatively and tried to find reasons for questions like why certain online resource have such high page ranking or have a good coverage?

The usefulness of the information source depends of what one is looking

for. Therefore I did not analyse the content itself - discussions, comments and websites - in order to find which one is better, the official or the unofficial source. Instead I examined the existence of different point of views and conflicts in free Internet. My intention was to reveal how different kind of views are competing in the free Internet and how those views become visible to the users of the Internet.

Fear and control are two keywords that are essential to my study. Fear of loosing control, fear of decontextualisation, fear of misuse. These fears are well documented in museum literary. I studied if there are any basis for these fears by analysing cases where cultural heritage materials are used freely in the Internet. I was after the term "misuse" and what that would mean in the context of digital cultural heritage. Is there a such thing as misuse of digital cultural heritage? Or is it just a way to discriminate other point of views?

During my study, I created technical proposals that demonstrate some solutions for documentation, openness and long-term preservation of digital heritage. With these applications, I have tried to find generic design principles that could be applied to open information systems and would, as such, be suitable for small heritage organisations or individual enthusiasts. The main parts of these topics are handled in separate conference papers later in this work.

In the final state of my work I have two views that I have named as Free Culture view and Open Data view. The first one is non-technical view that is related to questions like what is the role of the museums and how people are able participate in digital heritage. The Free Culture view includes also more practical issues like how are digital materials licensed and what kind of restrictions are set by copyright legislation, for example. The Open Data view includes the technical issues of open digital heritage and it concentrates on restrictions set by the technology used in the field of digital heritage and - in the other hand - also the possibilities provided by the technology. The word "open" refers open access in machine readable way. This view also includes long-term preservation issues. Long term preservation is a complicated issue and managing that is often beyond the resources of small museums. Therefore I found it crucial that long term preservation is part of design principles of heritage information systems.

Both of these views are needed when dealing with digital cultural heritage. Free Culture is the human-centric part of the open digital heritage while Open Data is the machine-centric part. In order to create tools for digital heritage, one must be aware of what are the demands and what kind of environment we are operating on. Although I am quite critical of heritage organisations placing unnecessary restrictions to their digital collections, I hope that I am able to provide some constructive ideas for implementing open information systems. To put it shortly, I am looking ways to improve the openness of digital cultural heritage. This means finding the road blocks by studying current practises and situation but it also means building bridges by creating prototypes for more open information systems.

2 ASPECTS OF OPEN DIGITAL HERITAGE

We can start examination of open digital heritage with the definition of digital heritage by UNESCO:

> Digital heritage is made up of computer-based materials of enduring value that should be kept for future generations. Digital heritage emanates from different communities, industries, sectors and regions. Not all digital materials are of enduring value, but those that are require active preservation approaches if continuity of digital heritage is to be maintained. (UNESCO, 2010)

UNESCO's definition of digital heritage is clear and easy to understand. We would like to preserve certain part of the digital content and we call that part our digital heritage. But if one removes words "digital" and "computer-based" from that definition, what remains is a definition of tangible cultural heritage. Is it true that digital heritage is like tangible heritage but only in digital form?

Like Fiona Cameron states, there has been only little critical reflexivity of the meaning of digital heritage. Definitions of digital heritage are linked to the past.

> Digital heritage according to the UNESCO charter is made on the premise of something to save and preserve rather than something that is created or built. ... In this sense the idea of digital heritage is a paradox, one which refers to newly created objects or media and also to discourses of loss. (Cameron, 2008)

The UNESCO's definition emphasises preservation. The digital heritage is something that must be protected from vanishing. But while preservation of tangible heritage means protecting, storing and locking down, it does not mean same things with digital heritage. Digital heritage can be copied, converted, analysed and combined infinitely. The nature of digital heritage is not just different from tangible heritage, it is radically different. The idea of putting preservation as a number one priority is inherited from the practises and ideology around tangible heritage: We must first ensure that artefacts are safe since they are unique items and therefore non-replaceable. After that we might think documentation,

accessibility and any active use of the heritage for education or tourism. But the problem occurs if the priority of preserving is seen as a primary task in the field of digital heritage. Then digital heritage appears as a black hole that soaks up digital materials but nothing ever comes out and the potential of digital heritage is never fulfilled.

Creating and managing digital heritage is a complex issue and several topics are related to studying of digital heritage. There are papers concerning technical implementations of interactive exhibition applications, automatic annotation of collection data, ontology construction, participatory experiments, database design, digitisation, virtual reconstruction and user interface design. There are more theoretical papers about museums role in digital age, originality and digitisation, information presentation, virtuality, communication, and contextualisation of collections. Several academic disciplines are related to digital heritage: museology, art history, ICT, information science and museum informatics. What follows is that it is difficult to build a "big picture" of digital heritage: managing digital heritage is quite more wider concept than what UNESCO's definition of digital heritage itself suggests.

This work concentrates mostly on one part of digital heritage: digitized cultural heritage materials and knowledge management around them. There are lot of activities around different kind of digitisation projects. Heritage organisations have huge quantities of digital files that can be called digital surrogates of tangible objects. This new heritage - a networked heritage - has different rules and there a more participants with different needs.

2.1 Digital surrogates

The product of digitising process is often called a digital surrogate. However, since the targets of digitisations vary greatly, not all digitisations are surrogates. For example, a digital scan of an photograph can be considered to be a digital surrogate of the original image since it preserves the main quality of the photograph; the surface texture. Or a digitisation of an analog audio recording is a surrogate because listener cannot distinguish the original from the digital version. But a digital photograph of a building serves hardly as a digital surrogate of that building because of the loss of spatial information. But an accurate 3D-scan of the building can serve as a surrogate since it preserves spatial information and surface texture information of the building. Abdelaziz Abid and David Arnold defines digital surrogates as follows:

> A significant part of digital heritage consists of the product of the digital reproduction of pre-existing works, which may consist of texts, images, and sounds, or which may be of an audiovisual, graphic, photographic or cinematographic nature. This digital 'double' does not claim to be an identical copy of the initial work, but only a representation of it. (Abid, 2007)
>
> The digital surrogate is the closest fidelity to the actual object that can be achieved digitally and theoretical representations for other purposes might

> be extracted from the surrogate. However, in practise it is unlikely that such levels of detail will be justified or achievable over all cultural artifacts and other categories may well be sufficient for identification or to get an impression in a Web page. (Arnold and Geser, 2008)p.63.

The process of digitisation is not just a simple analogue-digital transformation. Jim Lindner gives some examples of scanning of paper document:

> Some of the factors one could discuss would be the resolution of the sample or scan. What is the color temperature of the lamps during the scan? What is the amount and distribution of bits available to represent the color (black and white documents have color too and are there enough bits to represent the depth of the blacks as well as the distribution of the bits available throughout the color space)? Is the color space compressed in any way? Are there optics in the scanner and if so what is the distortion across the field (very few lenses are perfect)? What are the errors in registration? What is the linearity and sensitivity of the array.... you get the point. (Lindner, 2006)

Lindner uses a term "digital proxy" about digitisations that are not meant to be a surrogates - the closest fidelity to the original object - but that are good enough for viewing and referencing purposes. As an example Lindner mentions analogue video that is digitised and compressed with lossy codec. Lossy codec loses some part of the information but it makes file size smaller which in turn makes material easier to disseminate and use in different environments like mobile devices. Digital proxies can be used as a references. A small thumbnail image is sometimes enough in order to identify the painting, for example, for collection management purposes.

There are several interesting and challenging questions about digital surrogates. Why are digital surrogates created? What is exactly the content of a digital surrogate? If scratches and dust are removed from scanned photograph, is the end result a digital surrogate any more? Or is it an idealised version of the original object? And if a digital surrogate does not present the original object, how this relationship should be modelled in information systems? What does happen when digital surrogates are placed online? What happens when local turns to global?

Why are digital surrogates created?

It is possible to separate two different focuses in the process of creating digital surrogates: Preservation by digital surrogates and increasing accessibility by digital surrogates. These are not alternative focuses but it is important to see the difference. Preservation is part of the "serious" task of preserving cultural heritage by re-mediating it. If a physical item is lost then we have at least a digital version of it. The task is not to search new meanings or new perspectives concerning the digitised object but simply to preserve a digital copy of the cultural historical item. Therefore the focus of discussions about digital preservation concerns best practises of capturing all the aspects of originals and how to ensure the long term preservation of digital surrogates. If the digital surrogate is not shared

but instead it is only used in a part of the museum's internal preservation plan, it is merely a format conversion or re-mediation of artefact and in that sense it is an extension of tangible heritage. It is a local digital surrogate and nothing else. But when this surrogate is placed online, local turns into global.

Digital surrogates with global access frees us from physical limitations: We can access digital representations of physical objects from anywhere. On the other words, digitalisation can be used to increase accessibility of the materials. This, however, has side effects. Global access means that there will more participants involved. The meaning of the object, interpretation of it and the definition of the context are no longer exclusive right of the owner of the original object. In the global, networked heritage field there are more people that have something to say. These point of views are expressed in web sites, discussion forums, blogs and social media. From there they are picked up by search engines exposing them to the Internet users. In this context the questions are quite different from the preservation discussion. In what terms can be materials put online? What is the meaning of an object and who is able to define it? What is the role of the owner organisation in this process?

Although digital accessibility is a significant benefit of digitisation, there is an another benefit that is even more remarkable from information processing point of view; The essential benefit of digitisation is that it produces a digital file that can be copied, displayed, analysed, combined and modified with computer. Producing a digital file representing a physical is just the first step of digitisation and real benefits of digitisation comes visible after files are processed and used in combination of information systems.

Content and carriers

For the sake of simplicity we can start the examination of digital surrogates from digital scans of so called two dimensional objects like photographs, dia slides and paper sheets with text on them. Obviously these are three dimensional objects but the third dimension is usually bypassed in digitalisation. What happens when a photograph is scanned? The surface of the photograph is photographed by digital camera or scanner. As a result of this we have a representation of the visible surface of the object. This surface, in turn, is an analogue presentation of the original image that was captured by a film negative or a digital sensor. In the digitisation process we have captured one possible presentation of an original image, not the original image itself.

This situation becomes more clear when we introduce the term information carrier. We can say that the paper of the photograph is carrying the image and the image is actually an immaterial object. If we do this distinction, it helps us to operate on conceptual level of digital surrogates. The combination of immaterial image and its physical carrier forms an unique tangible item. The combination of the same immaterial image and another physical carrier forms another unique item. What links these two items is the image. The image that was captured by a photographer is an unique image, whether it was captured digitally or analog-

ically. This image is the element that links all following presentations - also those that are modified. The concept of the physical carrier is important when we create a data model for cultural heritage materials since we can separate information that concerns physical carriers from information that is carried by them. If different carriers are carrying the same information - like same image for example - then the carried image is documented only once which saves work and prevents duplicate data entries.

Value, usage and risks

John Unsworth has presented five questions and five answers that summarise several aspects of digital surrogates. Although Unsworth is mainly speaking from the librarian's point of view, most of these notions also applies to digital surrogates in general. Next I will present those questions and answers by Unsworth with my own comments.

> 1. When can a digital surrogate stand in for its source?
>
> When it answers to the needs of users. (Unsworth, 2004b)

What Unsworth means is that unless we don't need to examine the physical item, a digital surrogate is enough if we think that it is enough. We can compare paintings or study details of paintings if we have good digital surrogates of them and we are happy with the results. Otherwise we need better surrogates or access to the originals. So the question is about the needs of the users and the quality of surrogates. When these two meets, a digital surrogate stands for its source.

> 2. When can a digital surrogate replace its source?
>
> If the source is not rare. (Unsworth, 2004b)

Unsworth speaks about mainly library materials here. In the library context relative new newspapers are not rare sources and since they are probably archived for example by national archive, they can be replaced by digital surrogates. In the museum context the question is a little bit more complex. For example, children's school drawings in the museum's collection from present day are not rare. Still, every drawing is unique. Whether these drawings could be replaced with digital surrogates remains a difficult question.

> 3. When might a digital surrogate be superior to its source?
>
> In cases where remote or simultaneous access to the object is required, or when software provides tools that allow something more or different than physical examination. When the record of the digital surrogate finds its way into indexes and search engines that would never find the physical original. (Unsworth, 2004b)

Like I stated earlier, a digital version is machine processable which makes it superior to its source. For example, digitisation of a text sheet produces an image file that is useful for humans but for a computer it is just unsearchable, arbitrary data. But if the text is extracted from the document by optical character recognition, the situation changes. After that the content is accessible for computer and it

can be further processed by indexing or by searching for keywords, for example. Same applies also for other types of digital files: Computer vision application can be used to search persons from photographs, speech in video or audio can be extracted by voice recognition tools and 3D-models can be analysed for shape or measurements.

> 4. What is the cost of producing and maintaining digital surrogates?
>
> The cost of maintenance depends on frequency of use and the idiosyncracy of format, but beyond that it depends on technological, social, and institutional factors that are difficult or impossible to predict—which is an important reason for being cautious when one chooses to replace a physical object (the maintenance costs for which are known) with a digital surrogate (the maintenance costs for which are, to some extent, unknown). (Unsworth, 2004b)

Producing digital surrogates is a heavily labour-intensive work. Also planning and searching best methods takes time. This makes creation of digital surrogates quite costly. Still, the cost of a digitisation project can be quite well predicted. If scanning of n items takes t hours, then scanning n*10 items takes approximately t*10 hours, if planning is made well and materials are homogeneous. But maintaining digital surrogates is more complex issue. While digitisation is an one-time process, maintaining the materials is an ongoing process. The environment where maintaining takes place is changing constantly which makes costs hard to predict. If we are prepared to maintain materials infinite time, then we have infinite expenses.

> 5. What risks do digital surrogates pose?
>
> The principal risk posed by digital surrogates is the risk of disposing of an imperfectly represented original because one believes the digital surrogate to be a perfect substitute for it. Digital surrogates also pose the risk of providing a partial view (of an object) that seems to be complete, and the risk of decontextualization—the possibility that the digital surrogate will become detached from some context that is important to understanding what it is, and will be received and understood in the absence of that context. (Unsworth, 2004b)

It is easy to agree with Unsworth with the first two risks. Disposing the original when the digital surrogate is not complete is an unfortunate possibility that can happen. Directly related to this is the false impression of whole digital surrogate when it is in fact only partial. But the third risk needs more investigation. If one says that one of the risks of digital surrogates is the risk of decontextualisation then one also implies that there is a) a proper context and b) that it is the only (suitable) context. But if a digital file is taken out of the original context, then it moves to a new context. Therefore, the risk of decontextualisation can be rephrased as “the risk of re-contextualisation” or “the risk of re-interpretation”. Then we might ask if this is a risk or is it a possibility instead?

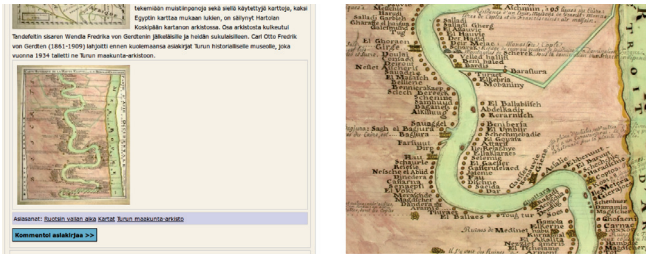


FIGURE 1 A reference image of the map of Egypt (left) and a close up from a larger version of the image (right). Images from extranet.narc.fi/arkistotkertovat/ by Arkistolaitos. Larger version is large enough in order to read the place names in the map.

Surrogate, proxy and reference

We can define three resolutions of digital surrogates. Resolution in this context does not mean directly technical resolution but resolution of the information. The first is the actual surrogate itself which is as precise as possible in given resources. The surrogate is impractical for everyday use and therefore a proxy is used. A digital proxy is a low quality representation of original object that is able to reveal the essential aspects of the original. In practise this means that for example the resolution of useful proxy images depends on the content of the image. If we have a letter, for example, then we should be able to read the text also from a digital proxy. Or if the original item is an architectural plan, the details of the drawing should be visible and architectural symbols should be understandable. In the case of musical notation notes should be recognisable. This means that proxy objects resolutions can not be the same regardless of the content but it is the content that defines which resolution is adequate. Otherwise digital file is just a reference, which is the lowest category of surrogates. The only value of reference is that it is able to point to original object. References can be used for example for collection management purposes. If references are used in online materials on their own, then their only purpose is that the museum can boast about the quantity of their online materials in the annual report without really providing any useful materials online.

Examples of both good proxy and more or less useless reference can be found in the online exhibition Arkistot Kertovat by Arkistolaitos (The National Archive Service of Finland). The figure 1 shows the map of Egypt. The map has place names written in it. The large version provided is large enough so that all the texts can be read. Figure 2 shows another document from the same online exhibition but from different archive. The image is a scan of the letter by Petter August Nummerlin. There is just one version of the image available. The text is unreadable and therefore the information value of the image is non-existent and therefore it is difficult to understand why this image is provided online. The image can be used only for identifying the original letter.

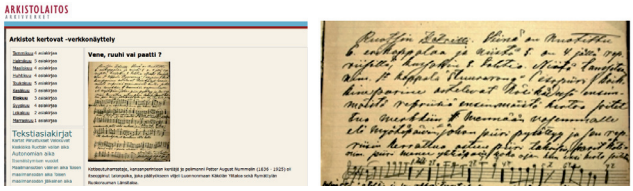


FIGURE 2 A reference image of document written by Petter August Nummelin from the archives of Turku via the online service Arkistot Kertovat. The text is unreadable due the low resolution of the image.

2.2 Born-digital content

An another widely used term in virtual heritage, besides digital surrogate, is born-digital content. This is the opposite of digital surrogate. Born-digital object is not analogue-to-digital transformation of tangible object but an object originally created in digital form like a website, a computer game or a digital text document. The variety of born-digital content is tremendous.

> Digital technology and the creation of ´born digital‘ content are indispensable aspects of cultural heritage management today. From low-tech documentation like Microsoft Office, html websites, PDF, and photography, to more complex technologies such as panoramas, object movies, laser/lidar scanning, scanning electron microscopy (SEM), x-ray fluorescence (XRF), Global Positioning System (GPS), 3D modelling, and distributed databases, to cutting edge techniques including Web 2.0, reflection transformation imaging (RTI), algorithmic generation of drawings from surface normals, and the family of photogrammetry influenced texture and 3D geometry acquisition tools, these new media types form a spectrum of opportunities and challenges to the preservation field that did not exist even 30 years ago. (Mudge et al., 2007)

Even if the museum considers itself as a “low-tech” museum, it is really hard to avoid born-digital content. Documents, the content of the databases and database management software are all born-digital objects. So are 3D-models, museum’s websites, info-kiosks and applications used in exhibitions. And if we think about documenting present-day then there is no way to avoid born-digital content both in the documentation and as an collection items. Therefore knowledge about born-digital content is essential part of digital heritage activities.

There are couple of issues related to born-digital object in museums. First there is the resource question. Museums have very varying chances to exploit the possibilities of using and producing born-digital content. Secondly, since born-digital objects contains information, there is a question of how these information sources - like databases - can be found and integrate by outsiders. And thirdly, the great variety of born-digital content forms a significant problem for long-term preservation. I will address these questions in more detail in “Search Engine Universe” and the “Long-term preservation” chapters.

2.3 Search Engine Universe

Search engines are the maps of the Internet. Currently Google is the market leader. 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.

One way to understand the search engine universe is to think about entry points or entry pages. The term entry page is used for the page that user lands on. In a regular site, every page can be an entry page. It means that user may land to any page of the site by clicking search result or link in email or some other site. The entry page might not be the page that the site owner wishes it to be. For example, when a museum introduces a new site for their image collection, they usually provide one entry point, namely the URL of the site's main page. This is the "main door" of the service and it is functioning as such as long as people a) know the address and b) remember the address. When people forget the address or the address is difficult to remember, they start to use search engines. People also might find the site accidentally during they searches if information of the site is exposed to search engines. If the site is a dynamic database, it might be possible that there really is only one entry page. Then the content of the site is completely invisible to search engines and the site falls in the deep web.

Search results in Google are objective in that sense that although it is possible to by adds that are attached to certain keyword, it is not possible to by better page ranking. The idea of "objective" page ranking was very important to Google (Auletta, 2010, p. 86 - 89). After all, it wanted to be the best search engine. If page ranks could be bought, then it would be just a paid catalogue of websites. So in this, anti-commercial way Google's search results are objective. But in the search engine algorithm level search results are not objective. Algorithms are made by humans and there is no way to determine objective method to rank pages since ranking factors can be selected multiple ways. Google uses its own PageRank algorithm, which - to put it simply - uses links between pages in order to define the importance of the page. In a way, PageRank is a popularity contest of web sites. This means that quality of the content is not necessarily the main factor in a good page rank. Something that is good to remember when speaking about page ranks.

However, because good ranking in search results is important for everyone who wants to be visible in the Internet, it is not surprising that different kind of ways to improve page ranking have been developed. Search Engine Optimization (SEO) is a term that means different kind of techniques that sites (especially commercial sites) are using in order to increase their ranking. Commercial SEO is usually well targeted and the effects can be measured. But there are also other kind of search engine manipulation that is more like a revenge or practical pranks. Google Bomb and Googlewashing are referring to practises "such as creating large numbers of links, that cause a web page to have a high ranking for searches on unrelated or off topic keyword phrases, often for comical or satirical

purposes"(Wikipedia, s. a.).

The most famous Google bomb case in Finland was targeted against excultural minister Tanja Karpela. A Google search with a term "turha lehmä" (useless cow in English) resulted the home page of the Karpela as a first hit. Internationally the best known Google bomb was when the term "miserable failure" resulted the biography of George Bush. These examples shows that manipulation of the search results is possible in certain extent. Still, it is almost impossible to affect what is showed in parallel in the result page. The Tipitii exhibition case later on this work shows one example of this. While the exhibition of broiler production history did not address any critical point of views, those views are dominant in search engine universe together with the name of the exhibition.

Search engines have a crucial role in everyday web usage. Therefore they have also major role in finding digital heritage materials. Search engines do not separate heritage institutions' sites from school projects or personal web pages. Although this can be seen as a problem that threats museum's authority, it can be also seen as a positive aspect that makes digital heritage multivoiced, interesting and vivid. In any case, search engine visibility is crucial if one wants to say something in the Internet. And with participatory web everybody can have their own voice.

2.4 Participatory Web

Many heritage organisations have earlier seen the Internet as a new broadcasting channel similar to television or radio where producers create information that can be accessed by the receivers. It has been a controlled way to distribute information to a wide audience. Diana Zorich uses the term "information delivery mechanism" in her article from 1997 which confirms the idea of one way communication.

> The Internet is an embarrassment of riches on both complimentary and pejorative sense. It offers tremendous potential as an information delivery mechanism, but it is chaotic environment that requires user to wade through its vast resources. (Zorich, 1997)

But the Internet has changed and the way the Internet is seen has been changed. There are more and more producers of information. Everyone can contribute to the pool of digital materials almost without restrictions. The traditional dichotomy in media structure with producers and receivers no longer applies. Like Mayer-Schönberger states, sharing is now more crucial than merely accessing:

> [...] before 2000 it [the Internet] was a tool to access information, today it is also a tool to share information. This implies a shift from passive recipient to active contributor, quite similar to what digitization has done for producing and consuming information itself. (Mayer-Schönberger, 2009, p.85)

This change that Mayer-Schönberger refers, means that the Internet users are not only users but participants. Although there has always been something rebellion,

or even anarchistic, in the nature of the Internet, this rebelliousness is now in the reach of many more. People do not have to have any technical expertise in order to have their views presented in the Internet.

The rise of social media has again altered the way the Internet is used. Manuel Castells uses a term "mass self-communication" when he speaks about new digital forms of communication:

> It is mass communication because it reaches potentially a global audience through the p2p networks and Internet connection. It is multimodal, as the digitization of content and advanced social software, often based on open source that can be downloaded free, allows the reformatting of almost any content in almost any form, increasingly distributed via wireless networks. And it is self-generated in content, self-directed in emission, and self-selected in reception by many that communicate with many.(Castells, 2007)

Castells states that it becomes increasingly difficult for governments to hide or manipulate information(Castells, 2007). The guardians of "correct information" have too many opponents, there are too many eyeballs watching and too many reporters reporting of wrongdoings. New communication forms had a crucial meaning in recent events in Egypt, Tunisia and elsewhere in Mid-East. The crowd was called to demonstrations via Twitter and Facebook and messaging about latest events was spread through the same channels.

> It is my contentation that the digital, networked age in one that can be, and is, amenable to just this kind this kind of horizontal, communicative action, and lend itself to a horizon of dissent, resistance and rebellion. (Hands, 2011)

The social media has undoubtedly changed things. Although it can be said that social media is open for rebellion or resistance, the content and platforms itself are not open. Facebook is closed system owned by a private company and so is Twitter. Most of the data inside Facebook is not openly available hence the private nature of the data. Facebook is an exclusive system. You have to give your personal information - including your phone number - to the Facebook in order to registrate. If you are not willing to do that you are not granted to access the data. The nature of social media is in present tense, "react now" type of participation. You do not find deep analysis of current topics in Facebook pages but you do find links to them as any other material outside Facebook. Massive amounts of information exists outside social media but the social media has created a new way to access that information.

Although social media has partly replaced the functionality of search engines by providing they own searches, search engines are still main entry points to the Internet. Search engines are trying to index all the content despite the organisational barriers. Facebook reveals some information to search engines but the most of the data is kept hidden. In order to have visibility in search engine universe one must create "traditional" web content. What has changed is the ease of creating and finding digital resources. Before users must have known at

least the basics of HTML and how to upload files in the server via FTP etc. Now starting one's own blog or publishing a video is few clicks a way.

The Internet in its current form allows anyone to make a statement but it must be done in a way that it can be found. Especially when you move to more specific areas that are outside of mainstream, everyone can have a voice and can be heard anonymously or publicly. This has effect also the ways digital heritage is created. Everyone can contribute without gatekeepers and results are picked up by search engines. This puts all heritage institutions to a new situation: they have lost their monopoly to "say how things really are", like governments have lost their ability to manipulate or hide information, 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.

2.5 Free software/Open source movement

If we want to define open digital heritage, then we must first study who are the participants of creating digital heritage and what does the term "open" mean in this context. The term open can be understood different ways. Open can mean accessibility, it can mean possibility to participate or it can mean total freedom to re-define or re-use materials provided. In the open source movement the term open is formally defined by the license framework and therefore it offers a good starting point for the openness in the digital world.

In his famous article The Cathedral and the Bazaar, Eric Raymond (Raymond, 1999) presents two alternative methods for software development. In the cathedral model software is build behind closed curtains by a nuclear team and only the final product is visible to end users. This is how proprietary software, like Microsoft Word or Adobe Photoshop, are created. The bazaar model, in other hand, is a noisy place where everyone can participate depending on their skills. Linux, Android, Firefox and OpenOffice are common names for most of us and they are living proofs - among thousands others open-source projects - of the power of the bazaar model. How are museums seen in the light of cathedral and bazaar models? Are museums cathedrals with their proprietary information products? Could museums benefit of the bazaar model - and contributions of hacker community - and what would that require from museums?

Steven Weber has collected the ways of how open source movement has been characterised. (Weber, 2000)

- A particular methodology for research and development
- The core of a new business model (free distribution of software means that new mechanism for compensation and profit need to be created)
- The social essence of a community, a defining nexus that binds together a group of people to create a common good.

- A new 'production structure' that is somehow unique or special to a 'knowledge economy' and will transcend or replace production structures of the industrial era.
- A political movement.

These formulations show that open source movement is not just about programming and source code. Actually it is misleading to speak about open source movement alone since there are two, quite different ideologies of this movement. The open source movement emphasis the practical benefits of open source software. The open source development model is simply good way to develop software. It "works" for business and it works for individuals in their own projects. On the other hand, the Free Software movement highlights ethical issues and freedom of information in general (Free Software Foundation, 2007b). Free software is important since it helps people to get access to information regardless of their wealth or social status. Free software is a way to empower people. From now on I will be using the term the Open Source movement in a sense that is also includes the Free Software movement.

Open source movement has been successful model for collaborative software creation. One of the reasons is that open source movement has a licensing framework that was designed for protecting user's rights to modify, use and distribute the software. Open source communities also offers various tasks to contributors depending their interests and skills, like translators, bug reporters, tutorial makers and even book authors, which increases the amount of potential contributors. And finally, maybe the most important reason in the success of free software is the lack of control. Developing, distribution and using can be started without any permissions or any bureaucracy.

Hacker Culture

Hacker culture is an essential part of free software and open source movements. The term "hacker" is in this context a positive term, that means a person who:

> [originally, someone who makes furniture with an axe]

1. A person who enjoys exploring the details of programmable systems and how to stretch their capabilities, as opposed to most users, who prefer to learn only the minimum necessary. RFC1392, the Internet Users' Glossary, usefully amplifies this as: A person who delights in having an intimate understanding of the internal workings of a system, computers and computer networks in particular.
2. One who programs enthusiastically (even obsessively) or who enjoys programming rather than just theorizing about programming.
3. A person capable of appreciating hack value.
4. A person who is good at programming quickly.
5. An expert at a particular program, or one who frequently does work using it or on it; as in 'a Unix hacker'. (Definitions 1 through 5 are correlated, and people who fit them congregate.)

6. An expert or enthusiast of any kind. One might be an astronomy hacker, for example.
7. One who enjoys the intellectual challenge of creatively overcoming or circumventing limitations.
8. [deprecated] A malicious meddler who tries to discover sensitive information by poking around. Hence password hacker, network hacker. The correct term for this sense is cracker. (Raymond, 2004)

Hacker culture highlights two freedoms; freedom of information and freedom from regulation. Hackers have fought against the use of Digital Rights Management (DRM), hackers have opened closed protocols, they have jail-breaked devices, like mobile phones, so that owners really have control of their devices and they have revealed numerous security issues that are threading average computer users. In additon, all kind of bureaucracy is very much disliked by hacker community like Pekka Himanen states in Hacker Ethic:

> In the hacker model, the individual simply starts creating, without any bureaucratic formalities, and passes her or his creation on to others directly without any complicated legalese (Himanen, 2001)(p.149).

Although hackers can be seen as a resource - skilled individuals that are willing to use their spare time for interesting projects - it is a resource that is not easy to tame and that is almost impossible to control as the examples of Xara and Kinect shows. The source code of vector drawing program called Xara was open-sourced and there were hopes that hackers would help to port Xara to Linux and Mac. The problem was that not all the code was not released but part of the drawing functions remained closed. Practically it meant that the application was not open-source program because without the drawing functions, the application was not functional at all. As a result there were very little contributions from external developers and open-sourcing Xara failed in the terms of gaining active hacker community. (Willis, 2007)

The Kinect case demonstrates the self-guidance nature of the hacker culture. Microsoft released a new kind of game controller called Kinect for its game-console Xbox in 2010. It was the first cheap consumer device that was able to do 3-dimensional motion tracking. As soon the device was published, it grasped the attention of the hacker community. Soon a bound was promised for makers of open source driver for the device(Adafruit Industries, 2010b). Although Microsoft protested against hacking attempts(Wilson, 2010), Kinect was soon reverse engineered and open-source driver was developed allowing Kinect to be run on Linux and Mac(Adafruit Industries, 2010a). This "opened" Kinect for different kind of experiments beyond it's original purpose of playing games. It has been used on robotics, 3D-scanning and several user-interface experiments like Minority Report style, gesture-based interface(Savov, 2010).

Despite the fact that the terms "hacking" and "hacker" are usually associated to the computer culture, these terms can be used also outside of this specific culture. As Eric S. Raymond writes:

> The hacker mind-set is not confined to this software-hacker culture. There are people who apply the hacker attitude to other things, like electronics or music — actually, you can find it at the highest levels of any science or art. Software hackers recognize these kindred spirits elsewhere and may call them 'hackers' too — and some claim that the hacker nature is really independent of the particular medium the hacker works in. (Raymond, 2001)

It could be argued that the term "hacker" means a mindset that is seeking novel ways of doing things in a given system by streching the rules of the system. In the traditional meaning of the term this environment or system would be an electronic system. Besides, we can also see heritage organisations, cultural heritage and even culture in general as a system with certain rules, restrictions and conventions. If we widen the term "hacking" this way so that it includes also creating new meanings and new contexts by using existing cultural materials, then we are close to the term "cultural hacking". When someone uses cultural item in one's homepage or in a self-made Youtube video, for example, one simultaenously contributes to digital heritage by offering new context, new information or different point of view. This is the very essence of the participatory web: people use something existing in order to create something new.

Motivations and licences

If we are going to apply practises of open source communities to digital heritage, we must first understand why hackers - and other contributors - are willing to contribute to open-source projects. Lerner and Tirole have suggested that signalling is the key concept of understanding programmers willingness to use their spare time for open source projects (Lerner and Tirole, 2001). Signalling is an extrinsic motivation and it explains that by contributing to open source project a programmer signals one's expertise to other hackers and possible employers. Signalling theorem sees an open source community as a ranking system. Signalling theorem has been criticised for not taking into account other aspects of open source programming like fun or "scratching the itch". Many hackers have emphasised the importance of fun factor in programming. Programming is intellectually challenging and fun. This still does not explain why hackers publish they work for everyone to use. Bitzer, Schretti and Schröder formulate this as following:

> There are thousands of programmers who program new software because they need it and have fun programming it, but these programmers decide to earn money with the final product or are not willing to publish the source codes, but rather keep them as private software solutions. Thus, what is the motive to turn one's efforts into a public good, namely publish the source code of one's program and have it licensed under the GPL? (Bitzer et al., 2005)

When it comes to commercialising software, there are couple reasons that might affect developer's decision about the future of the software. In order to commercialise the software, you need first to make a software that has commercial

potential. Then one need to organise money transactions, provide support services and find out the legal and organisational issues of running business. All this combined may be not as fun than just programming and letting people to use the product freely. And there is still one motivation that was not covered in earlier motivation explanations: the need of help. When software project grows, the workload also grows. There are more bug reports, questions, more demands and documentation to do. There could be also some areas on programming that are unfamiliar to the developer and the developer has no time or interest to learn that specific area. At this point the developer needs help. If the source code is closed and there are no resources to pay for development, then it could be very difficult to find volunteer developers. Open sourcing might then be only way to solve the situation.

It is also good to point out one of the aspects of open source software that is linked to motivation. It is the licensing framework that is designed to open source projects. This licensing framework guarantees that it is safe to contribute to an open source project. Although everyone can download and modify and redistribute my code, no one cannot take it from me. The code is still mine and it can not be locked in.

Open source licences are an essential part of open source movement. There are several different kind of licences like GPL, Apache, BSD and MIT, that define what rights and obligations users have. Maybe the most important licence is the GNU General Public Licence or GPL(Free Software Foundation, 2007a). It was written by Richard Stallman and it is the most pure Free Software licence. GPL guarantees that the source code is never closed since users can not apply additional restrictions to the code.

> In essence, anyone is free to do anything with the GPL code except things that restrict the freedom of others to enjoy the same freedoms. In practise this means that a program derived from GPL code must also be released under the GPL with its source code. (Weber, 2004)

The GPL licence is designed especially for free software. If someone is using his time and energy for an open source project, it is the GPL that makes sure that the contribution is kept safe. The contributions benefit all but no one can omit it or build a closed source product top of it. On the other hand, there is nothing that prevents modifications of contributions. If there are flaws in the code, they can be corrected by other volunteers. This way GPL offers a maximum support for collaborative work of volunteers by protecting the work done but allowing improvements without permissions from original authors.

Best practises of open source communities

There are at least two possible scenarios when thinking about digital heritage organisations and open source culture. The first point of view is how heritage institutions could directly benefit of hacker culture. It is a fact that museums - for example - have quite limited resources on ICT. The question is what museum should do in order to lure hackers to “work for them”. On the other hand it is

possible to study of what kind of practises open source communities are using and how those can be applied to digital heritage in order to make it more open. This view concentrates of searching best practises of open source communities and examining if those could be used in heritage institutions.

I stated earlier that open source methodology of creating software has been successful and the idea of open source can be found also outside software development. From the digital heritage's point of view interesting questions are why it has been successful and can those elements to be transferred to the field of digital heritage.

> Open-source methodology has already spread well beyond software development: In the world at large, the Human Genome Project is a famous example. Over the coming decade we're certain to see this new mode of production locked in mortal combat with older methods and the legal and ideological commitments that they entail. It will be interesting to see whether, at this critical juncture, the university comes down on the side of freely shared ideas. (Unsworth, 2004a)

Like John Unsworth notes above, the ideology and practices of open source movement does not concern just software anymore. Openness is a model for open data movement, open education movement, open hardware projects, open access practices and open government. These all see the value of sharing and participating as a driving force for gaining something new.

It would be tempting to argue that the one of the main reasons for the success of free software is the lack of control. All kind of bureaucracy and permission asking is very much disliked by hacker community like Pekka Himanen states in Hacker Ethic:

> In the hacker model, the individual simply starts creating, without any bureaucratic formalities, and passes her or his creation on to others directly without any complicated legalese. (Himanen, 2001)

However, there are control structures inside free software communities. Therefore it is more correct to say that the key for the success of free software communities is not the lack of control but the way control is divided and the way one can enter into the control system. An individual programmer is free to work with free software in her/his private space. That means that programmer can download the code, tamper with it, redistribute the altered code and so on. But when the programmer enters to a community space, then the programmer also enters to a control system. The control system defines what are the goals of the project, how things are organised, which are areas of responsibilities, what coding practises are used and so on. Steven Weber presents important points related to control:

> The popular image of open source as a bazaar does capture the feeling of an ideal type. It is an evocative image. But it is analytically misleading and it is best to drop it. The key element of the open source process, as an ideal type, is voluntary participation and voluntary selection of tasks. Anyone can join an open source project, and anyone can leave at any time. (Weber, 2004)

The very essential part of the success of open source is the fine balance between freedom and control. In that sense, the bazaar metaphor might give a wrong idea of the role of control in the open source community. It is the freedom that invites people to participate but it is the the control that makes it possible to run projects successfully. As I mentioned earlier, the legal framework is important part of this balance. It protects the work of developers still allowing others to participate.

Open source projects are not democratic communities. On the contrary, there is a small group of developers that makes decisions about the future of the project. This group is not a representatives of the users of the software but merely a self-selected group that rules by merit. This model can be called meritocracy. For example the Apache project explicitly states that meritocracy is how the project works:

> As the group started to develop their own version of the software, moving away from the NCSA version, more people were attracted and started to help out, first by sending little patches, or suggestions, or replying to email on the mail list, later by more important contributions.
>
> When the group felt that the person had "earned" the merit to be part of the development community, they granted direct access to the code repository, thus increasing the group and increasing the ability of the group to develop the program, and to maintain and develop it more effectively.
>
> We call this basic principle "meritocracy": literally, government by merit. (Apache Software Foundation, 2011)

Meritocracy is the career path in the open source world. Degree, location, race or age is not a factor. Only thing that means is what one can do and what one is willing to do. So, even though the open source world does not follow democratic ideals, it still promotes equal opportunities.

The museum culture seems to be quite opposite when compared to the culture of open source movement: The legal framework is protecting copyright holders, participation of outsiders is usually limited to one, pre-defined role and permission must be asked for using, modifying and distribution. Despite this conflict between these two cultures, is there a possibility to create open heritage movement based on the best practises and ideology of open source movement?

2.6 Machine Access

Application Programming Interface (API) means that some service can be accessed programmatically. This makes it possible to built services or applications that uses and combines information from several sources without any human intervention. Relatively little noticed change in the Internet during recent years is the appearance of free and public digital interfaces, more commonly known

programmable web. Google, Yahoo, Twitter, Flicker, and recently also some museums, are providing digital interfaces to their data. In addition, there are almost three thousand APIs listed in programmableweb.com and hundreds of mashups and applications are using these interfaces.

When information is copied, it loses its connection to the original source. Therefore any updates or changes are not visible in the copy and vice versa. Only way to keep information connected to the original source, is to provide some kind of machine access to the original source. Using an API requires at least some technical expertise and therefore digital interfaces are not directly useful for end-users. There must be a mediator - lets call them data hackers or data journalists - that programmatically builds way to use the data resource in the form of, for example, a mashup or mobile application. Data hackers are a special group that have the power to "re-mediate" information resources to others. For end-users, services and applications created by data hackers means more alternatives; different ways to use information, new ways to participate on knowledge creation, localisation of otherwise global services or just new ways of entertainment. This breaks the traditional structures between services and clients and it raises questions of authority, control and the role of participants. When a heritage organisation opens a public, digital interface, it is an invitation to create new uses of material. The use of API makes it easier to use and find data if the API is well designed.

But from the open digital heritage perspective APIs have certain issues also. First, it is possible that API requires a registration. In that case the API can be seen also as a control mechanism. Registered users can be tracked down and the provider of the API can punish the users that are somehow "misbehaving". The problem is that it is in the power of the API provider to decide what is considered as a misbehaviour. It can be excessive use of bandwidth or it can be using the data against the API provider's ideology. The second - a more technical problem - is that if the API is changed, then it will affect all applications using it. This is a common problem with any API. When API is developed further there will be a point where it becomes backward incompatible. This means that mashups and other third party compilations must be maintained actively. The third problem in the heritage sector with APIs is that there is no generic API that could be used. This means that even though museums started to provide APIs, in order to use them one should know details of every individual API.

However, APIs are not the only way to provide machine accessible information. Information can be stored in files and place the files online. Since the file formats can be anything, this approach has its own problems. But if file formats are open - for example XML-based formats - and files themselves are not subject of change, then file-based access may provide more stable, and more open way to information access. This is the approach that many Open data providers have used.

2.7 Open Data

Open data is an interesting phenomena from the digital heritage point of view. The idea is that data (that is usually created by publicly funded organisation) is made freely available. Being available means in this case more than just access. It means also right to distribute and reuse data. Open data can make government more transparent and it could help taxpayers to see how their money is spent. It can enable new scientific discoveries or it can reveal the socio-economical status of different districts. Open Knowledge Foundation defines requirements for Open Data as follows:

- Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably downloading over the Internet. The data must also be available in a convenient and modifiable form.
- Reuse and Redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other data sets.
- Universal Participation: everyone must be able to use, reuse redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use or restrictions of use for certain purposes (e.g. only in education) are not allowed. (Open Knowledge Foundation, 2011)

Raw open data can be scientific data, statistical information or governmental data like meeting minutes. Scientific data or statistical information is quite difficult or impossible to interpret without certain expertise. Not to mention of combining it with other resources which require also programming expertise. Also government documents requires usually some kind of background information. So open data is not aimed directly for large audience but it is aimed for information professionals or information hackers who are able to combine and interpret complex statistical data. A non-profit organisation Open Data Foundation formulates situation as follows:

> Data has many sources, the administration of surveys and the monitoring of transactional flows and registers being some of the most common. In order to become useful for the end-user communities, raw data commonly go through various editing, aggregation and analytical stages. While researchers and academics may find the micro-data useful, policy and decision makers and the and the general public are more commonly interested in the easier to manage high-level aggregates. Despite the existence of tools and the emergence of open metadata specifications, it is often not possible to connect the different parts of this information chain together. Such connection, however, is critical in fully understanding the data. (Open Data Foundation, 2011)

Digital heritage, in turn, is usually much more easily accessible for large audience. The nature of open digital heritage is more visual. Visual heritage is images

and information about images and images are something that is easy to understand also for public. The point is that pure heritage data is not very useful or even interesting without other materials related to data. The open database of paintings without images of the paintings is quite worthless for an average online users who wants to see the paintings. But just placing materials online does not mean that they are really accessible. One must be aware of how people find information and how they want to use it. Mike Ellis gives an example of Open Data approach:

> Philosophically, Open Data becomes about the value that you want to realize through your assets. Consider two possible scenarios: one where your organisation publishes a PDF on your website describing an important object in your collection, and one where you instead add this same information to Wikipedia as HTML under Creative Commons license of some kind. The latter approach is always going to get more people finding, reading and using the information you've written than the former - this is the power of a huge, globally recognized brand like Wikipedia coupled with more open licensing. (Ellis, 2011, p.181)

Another difference between open data in general and digital cultural heritage is authority. If we look the requirements of open data presented by opendatamanual.org, we see that it is does not mention modifying. This makes sense since the whole idea of open data is that we have a trusted source of information. Modifying the data would render the whole approach invalid. In open data the authority of the data is not questioned. In open digital heritage the situation is a little bit more complex. While governmental organisation provides open data, it is normally just data that is not copyrighted. But in heritage organisation there are also images, sounds, videos or even 3D-models. For example, a typical case in museum field is that there is an online collection record with data and image or images. There is an information part and a materials part in digital heritage. Even though the example of Ellis above is illustrative, it is not totally correct. Ellis seems to forget that once information is written to Wikipedia, it becomes editable by anyone. So it is not necessarily "information you've written" but information that someone has written and that someone has possibly edited. The authority of the museum does not apply to Wikipedia.

By using the requirements of the Open Data, we can formulate requirements for open digital heritage materials. There are three conditions that must be met so that we could say heritage materials are open.

- materials can be downloaded and accessed (technical openness)
- materials can be found (accessibility, search engine visibility)
- materials can be re-used (licensing, moral openness)

The requirements of open data and open digital heritage are very similar in every aspect except when it comes to commercial use. According to the Open Knowledge Foundation there must be no restrictions about commercial use of the data. This is because the one idea of Open Data is to create business opportunities with open data. That, however, might be problematic for digital heritage. Although

collection data could be used for commercial purposes there might be limitations for images, videos and audio materials.

2.8 Free Culture and Creative Commons

A third "open movement" after open source and open data movements - although I am not sure if this is a right time order - is the free culture movement. The term free culture is an opposite of "permission culture" that is enforced by copyright laws and currently intellectual property legislation. The term itself was made popular by Lawrence Lessig with his book Free Culture.

> A free culture supports and protects creators and innovators. It does this directly by granting intellectual property rights. But it does so indirectly by limiting the reach of those rights, to guarantee that follow-on creators and innovators remain as free as possible from the control of the past. A free culture is not a culture without property, just as a free market is not a market in which everything is free. The opposite of a free culture is a "permission culture"—a culture in which creators get to create only with the permission of the powerful, or of creators from the past. (Lessig, 2004)

The roots of the free culture are in the open source movement as Lessig describes in his book.

> The inspiration for the title and for much of the argument of this book comes from the work of Richard Stallman and the Free Software Foundation. Indeed, as I reread Stallman's own work, especially the essays in Free Software,Free Society, I realize that all of the theoretical insights I develop here are insights Stallman described decades ago. One could thus well argue that this work is "merely" derivative.
>
> I accept that criticism, if indeed it is a criticism. The work of a lawyer is always derivative, and I mean to do nothing more in this book than to remind a culture about a tradition that has always been its own. Like Stallman, I defend that tradition on the basis of values. Like Stallman, I believe those are the values of freedom. And like Stallman, I believe those are values of our past that will need to be defended in our future. A free culture has been our past, but it will only be our future if we change the path we are on right now.

And like Free Software does not exclude commercial activities outside of the movement, so does not Free Culture deny property of creators of cultural works. The question is about a balance between the rights of the users and the rights of the creators.

> Like Stallman's arguments for free software, an argument for free culture stumbles on a confusion that is hard to avoid, and even harder to understand. A free culture is not a culture without property; it is not a culture in which artists don't get paid. A culture without property, or in which creators can't get paid, is anarchy, not freedom. Anarchy is not what I advance here. Instead, the free culture that I defend in this book is a balance between

> anarchy and control. A free culture, like a free market, is filled with property. It is filled with rules of property and contract that get enforced by the state. But just as a free market is perverted if its property becomes feudal, so too can a free culture be queered by extremism in the property rights that define it. That is what I fear about our culture today. It is against that extremism that this book is written. (Lessig, 2004)

In the book Free Culture Lessig heavily criticises current copyright laws and how they are pushed further and further in a digital age in the expense of the rights of the users. As an example he presents his ebook reader software by Adobe. The reader software allows to set different kind of permissions for the content. For example, with one title, a user is able to copy ten text selections to the clipboard in every ten days. Or user can print only ten pages every ten days. The permissions also dictates if the user is able to listen the book with voice syntheziser. These absurd limitations are only possible in a digital environment and they go beyond the copyright laws.

> When my e-book of Middlemarch says I have the permission to copy only ten text selections into the memory every ten days, what that really means is that the eBook Reader has enabled the publisher to control how I use the book on my computer, far beyond the control that the law would enable. (Lessig, 2004, p.151)

Lessig was one of the founders of a non-profit organisation called Creative Commons. The organisation "develops, supports, and stewards legal and technical infrastructure that maximizes digital creativity, sharing, and innovation". They have created Creative Commons licence (CC), which can be used define what rights a creator wants to keep and what rights can be given to users. With Creative Commons one can define whether a work can be altered or not, and a permission for commercial or non-commercial use of the work can be granted.

Creative Commons license has been influential and it has been widely accepted as the free license for the web: Flickr has its own Commons section called Flickr Commons, Google has a CC search in advanced search options and Wikipedia uses CC as a default media license. The success of CC has shown that functional license framework for sharing digital objects is indeed possible.

2.9 The role of technology in Digital Heritage

In his book *Software Takes Command* Lev Manovich criticises new media theorists, scientists and culture critics for their lack of interest of software that enables phenomena they are studying.

> "Software is still invisible to most academics, artists, and cultural professionals interested in IT and its cultural and social effects." (Manovich, 2008, p. 4)

One might argue that there is very little difference if a literature academic is aware of technology needed for printing books or not. They study the content that is

carried by certain layout using certain font printed in certain kind of paper. Although the presentation (layout, fonts, paper quality, size) of the book affects the reader's experience, it does not matter when we discuss about book's structure, plot or characters' motivations. A book itself is established form of presentation unlike digital media.

The role of technology in digital media is more crucial. Digital media is accessed through software and hardware layers that are constantly changing. We do not experience files and documents as they are stored but we see them through applications. And way we experience them is directly controlled by software: Viewing photos in home PC with image presentation application, listening music from an internet radio or reading a blog that has embedded Youtube video. Manovich speaks about software culture:

> We live in a software culture - that is, a culture where the production, distribution, and reception of most content - and increasingly, experiences - is mediated by software. (Manovich, 2008, p.19)

In practical level this software culture can seen by how technology sets several border lines for digital content. First of all, it has a major role on accessibility. Some examples:

- Certain file formats require certain software efficiently excluding or including users.
- Using non-free codecs in video or audio files requires users to have specific player software.
- Providing an application that is not a cross-platform application (i.e. application that works on different operating systems like Linux, Mac and Windows) is again a selective criteria.
- The bandwidth (which is a resource but can be seen also as technical limitation) defines how much data user can receive and thus effects of using of, for example, video services.
- Bad web design (using non-standard, browser specific features) might render web pages totally unusable when accessed certain web-browsers (including mobile phone browsers) preventing users of those browsers to access to the page.
- A website might be not indexed by search engines because the site is created "on the fly" from the database and therefore the content of the site stays in hidden net.

Let's say an academic is studying accessibility of a museum's website without knowing restrictions mentioned above. He has used only Microsoft's Windows operating system with its default browser Internet Explorer. He notices no technical problems when accessing site that is (badly) designed to work only with that combination of an operating system and a browser. Then it does not matter how detailed analyses he makes about navigation structure or cognitive load of the site if he fails to understand that site's design makes it inaccessible for users with other browsers and operating systems. Chosen technology is exclusive by

its design and that affects the accessibility. If this point of view is missing, then the whole study of accessibility lacks of one major factor; the accessibility of the used technology itself.

Or if a museum exports their collection database to the web without understanding that their dynamic site is not indexed by search engines. The museum can then say that their collection data is on the web which is only partially true. Yes, the database is online but is it really accessible? If it is not found by means that average users uses when they search information from the web, then it does not exist because it cannot be found.

Secondly, technology has affect on what can be done. For example, a museum's collection management software defines what kind of information can be added to the collection items and how that information is organised. The question is how much power individual museum had in the design process of that software? My point here is that technical issues cannot be ignored when studying digital heritage. This is the challenge of digital heritage: Digital heritage is heavily related to ICT but the substance of digital heritage is heavily related to humanistic field. These two cannot be handled as a separate entities nor is the other more important than the other. Changes in technology affects of how content can be created, processed and disseminated and the cultural context of the content affects of what can be achieved with technology and what is desirable.

Tomislav Sola, professor of museology, has discussed the role of information technology in museums and museum profession in his essay collection *Essays on museums and their theory : towards the cybernetic museum*. Sola uses the term IT but more up-to-date term would be ICT (Information and Communication Technology). The word "communication" is essential to describe information technology today. ICT is about communication: human to human, human to computer and computer to computer communication. Sola sees the role of ICT as a positive, almost ground breaking force that will affect museums and their way to do their work.

> The nature and the logic of IT apply so well to the entirety of the cultural heritage concept and the totality of heritage that its influence will have crucial consequences upon its further development. [...]
>
> [...] Community and cohesion may just be a way to approach the challenge of the age of synthesis we are entering. It is quite curious that IT reinforces the point more clearly than any theoretical or practical proposal has so far. Synthesis is more than cross-reference and a practical tool for research of meaningful presentation. It is the way out of the labyrinth of specialized area of knowledge and an end to the incessant manipulation of laymen's interests. (Sola, 1997)

But this development has major roadblocks that are partly linked to the whole profession of museum workers and their understanding of their own work.

> The new technologies are still much too incomprehensible to the majority of professionals, as they have trouble of going beyond the form and method. One should not blame producers of IT for not being able to induce enough understanding as they rarely know the subtleties of the museum mission.

> The profession itself has difficulties with that same understanding, thus endangering its own survival in the face of rising competition. (Sola, 1997)

The problem is in the museum professionals that do not know the possibilities and functions of ICT. Even more serious problems is - according to Sola - that the museum profession has no clear understanding of their own profession. The other half of the problem is ICT providers, that are not familiar with with the "subtleties of museum mission". If we combine these problems, then we have a situation where a museum, which does not understand ICT and that does not know what they need, is acquiring ICT solutions from a provider that knows ICT but do not understand the needs of the museum. Then, one might ask how anything good could come out of that.

Sola seems to think that ICT has huge potential for cultural heritage. But ICT might also be a new barrier for cultural heritage. Instead of creating new "mega-profession" - the term Sola uses - ICT might be a restrictive force that taints the development of a heritage institution. ICT can be a black hole that vaporises all resources and data thrown in it. Rigid information systems may freeze the development of documentation practices for good. Formalisation, semantic techniques and information integration can be a burden as well as they can be a blessing. Strict formalisation, while serving well engineering of ICT, might be totally unsuitable for contextually rich documentation purposes. Semantic techniques can be cause more confusement than clarity and overhelming information integration can lead to an unsuitable and limited data model. It all depends on a viewing angle and balance. Sometimes the best practices of ICT might conflict with the best practices of cultural historical field. Digital divide between resourceful and less resourceful institutions is also an issue.

Although there are some serious roadblocks on the way to the new heritage, I still agree with Sola that eventually ICT will change the role and profession in museums. Partly that has already happened but the major change is happening because the pressure and competition coming from open networks. There is a pressure to open heritage materials and while they are exposed to the Internet, these materials are confronting new environment that does not recognise heritage institutions authority. They must deserve they place by other means.

3 PROBLEMS OF OPEN DIGITAL HERITAGE

Projects like Project Guttenberg and Mutopia Project make public domain books and notes available for everyone, Google Books has millions of books online, Wikipedia is often number one in the results for a heritage related queries, individual users and groups have their own online archives and museums that can be the best resources available on certain topic, best images of ancient coins can be found on online market places selling the coins. And all this information can be found with generic search engines. In addition there are terms like open source software, open data, open education, open access, open knowledge and even open hardware(Wikipedia, s. a.) marketing this trend of openness. But are there any signs of open heritage? Museums explicitly deny any use of their materials, the content of the museum's online databases can not be found with search engines, users can take part of the museum's participation projects but the ways of participation and the content created is strictly controlled by the museum and not even photographing is allowed in exhibitions. Did heritage organisations miss the "openness" trend?

For example, if we look UNESCO's home page, we will found encouraging words like this:

> World Heritage sites belong to all the peoples of the world, irrespective of the territory on which they are located.
>
> – http://whc.unesco.org/en/about/ accessed: december 2011

When we look at the Term and Policies section of the same page, we'll find Images disclaimer stating:

> All photographs and other elements of the UNESCO World Heritage Centre website are protected by international treaties on intellectual property and other applicable laws. The elements it displays may not be copied or retransmitted by any means without explicit authorisation from the World Heritage Centre. All photographs shown on this site are the exclusive or non-exclusive property of UNESCO or of third parties, copyright holders, who authorize the UNESCO World Heritage Centre to display and distribute them. Any use of photographs requires a separate request for authorization from the World Heritage Centre (email: wh-info@unesco.org).

> This permission may not under any circumstances be transferred to a third party.
>
> – http://whc.unesco.org/en/disclaimer/ accessed: december 2011

There seems to be a contradiction in the messages provided by the UNESCO's site. On the the other hand the website is clearly stating that heritage sites belong to all. But at the same time the website states that images about heritage sites are property of UNESCO or third parties and copying and redistribution of them is strictly forbidden. So heritage sites in digital form do not belong to all but only for copyright holders of the images.

We might ask if we need open digital heritage in first place. Why should museums - or heritage institutions in general - let people to use their materials without any payment? Aren't museums entitled to "own" their collections and therefore have right to define the destiny of digital surrogates of their collection? Similarly we might ask if a publicly funded museum really own its collection? If it was funded by tax payers money, then isn't it right that the public has free access to them? Stuart Hall discusses about British Heritage and construction - and constructors - of the Heritage in the article called *Whose Heritage? Un-settling the 'Heritage', Re-imagining the Post-Nation*. Although he is merely speaking about British Heritage and changes in its content and how the Heritage is constructed, much of the discussion also applies any western heritage. Hall states that (British) Heritage has confronted two challenges.

> Despite all this, the idea of Heritage has had to respond to at least two major challenges. The first we call the democratization process. Increasingly, the lives, artefacts, houses, workplaces, tools, customs and oral memories of ordinary everyday British folk have slowly taken their subordinate place amongside the hegemonic presence of the great and the good. [...]
>
> [...] The second 'revolution' arises from the critique of the Enlightment ideal of dispassionate universal knowledge, which drove and inspired so much of Heritage activity in the past. This has to be coupled with a rising cultural relativism which is part of the growing decentering of the West and western-oriented or Eurocentric grand-narratives.(Hall, 2002)

If we accept Hall's idea of democratisation process of heritage, then we can easily find justification for open digital heritage. In the first phase ordinary people were accepted as a part of heritage. Open digital heritage is the second round in the democratisation process of our cultural heritage. In this second phase ordinary people are able to interpret and spread heritage in digital form. This process is possible only by the means of open digital materials that people can share and re-use. Museums are not able do this democratisation behalf of the people. Despite the efforts of the museums, there is limited number of museum personnel and limited number of perspectives inside the museums. Not even crowdsourcing or user-generated content are not enough because of the limited freedom of those approaches.

Hall's point of the ideal dispassionate universal knowledge and its role in heritage activities can be accepted also from the museums' point of view. Collection documentation of museums is unemotional, clinical and in many cases also

trivial. Although this serves purpose in the internal collection management, it serves badly when museum wants to communicate with outsiders. On the other hand, relativism is a "built-in feature" of the Internet. All kinds of views - no matter how odd they seem to be - have their place in the Internet. This relativism becomes almost automatically part of online heritage materials if they can be re-used. This, and only this, allows people to participate in the heritage activity with their own terms.

In the beginning of this chapter I criticised museums for missing the openness trend. Still, a lot of things are happening in a more open direction also in the museum field. The Brooklyn Museum has licensed their collection images under Creative Commons and provides open application interface for fetching data to the user's own applications. Smithsonian Institution is in the process of creating Smithsonian Commons for helping users to re-use their materials more easily. Rijksmuseum has recently published an application interface that can be used for accessing over 110 000 public domain images. In the field of archaeology there are attempts to open data sets for public use. Many museums have published their images in Flickr Commons where people can download them for further usage. Is the museum field divided to open and not-open museums? What are the reasons for museums to open or not to open their materials?

3.1 Museums

Museums that have been responsible of collecting and preserving our tangible cultural heritage, are now in a new position. New demands have arisen concerning the role of museums in the society. Museums should be accessible for very different kind of groups of people, they should provide multiple perspectives for cultural historical material and they should be present in global networks. In the networked world, heritage organisations are competing with every other information resource there exists. Suddenly, the museum's collection can be a restriction that non-institutional online resources does not have. Heritage institutions must decide their role in a new environment which is essentially global, highly technical and under constant change.

An US-based survey by the Institute of Museum and Library Services (IMLS) lists several reasons why heritage organisations create controlled online collections instead of open online collections. The term "controlled online collection" means an online collection that applies some sort of restrictions to the re-use of collections.

> 1. Policy statements defining rights associated with digital collections are still nascent. [...]
>
> 2. Cultural institutions control collections in order to avoid misrepresentation, mislabeling or misuse of cultural objects. [...]
>
> 3. Non-attribution is also a major concern for cultural institutions as well as unauthorized commercial publication. (Eschenfelder, 2009)

Reasons can be categorised loosely to rights and fears. Problems with rights can be related to missing rights, unknown rights or rights that are restricting availability of the materials. The second category is the fears. According to this view control is needed in order to prevent misrepresentation or mislabeling. But the study also revealed that when cultural institutions creates open collections, the main reason for that is that organisations find open collections more effective:

> The primary reason why cultural institutions don't create controlled collection is the belief that open collections have a bigger impact. (Eschenfelder, 2009)

In this chapter I study some factors that are affecting of how museums operate in digital environments. These aspects consist of rights, resources, documentation practises and models, attitudes and changes in the museums' operational environment.

Collection management

> Even before the explosion of digital technology and information in the 1990s, museum leaders have clearly understood that museums are in the business of information. (Peacock, 2008)

As Darren Peacock states above, information handling is essential for museums. Most of this information is naturally connected to a collection management. Using computers for collection management in museums have a long tradition. First attempts were made in the early 1960s with mainframe computers. In the 80's, desktop computers and more choices on software made it possible for museums to have their own, computerised collection management software. (Jones-Garmil, 1997) Since then, smaller and smaller museums have been able to use some kind of computer-based collection management system. The base unit of these systems is usually a collection item or group of collection items. This item-centricity has long roots in museums and a very strong position in documentation practises as Fiona Cameron states:

> Item-level documentation of museum collections has, until recently, been taken for granted as one of the unchanging givens of museums practise. [...] However, the basic descriptive criteria such as "data of manufacture", "maker", and physical measurements, used by collection managers and curators in documenting museum objects, have remained largely unaltered. These formulaic descriptions are rooted in the long-established practises of curatioral disciplines such as art history, decorative arts, history, science and technology. (Cameron, 2007, p.166)

The information systems in museums are mostly designed from this item-centric (or object-centric) and internal point of view. The task has been to fulfil the needs of the museum. And those needs are closely related to a management of tangible collection items. Therefore it is not surprising that the base element of information handling in museums is a physical collection item. As Darren Peacock formulates this:

> Traditionally, museum information management has focused on supporting key collections operations: cataloguing, documentation, condition and treatment reporting, movements and curatorial research. Our information systems have been designed primarily to manage our collections as physical assets rather than to manage the information they represent or embody. (Peacock et al., 2004)

In a way, the information stored in these systems is a side effect. In other words, museum information systems are not designed to handle information as it is but only information that is strictly bundled to physical items. This leads inevitably problems when the scope of the documentation is broadened beyond tangible objects. In fact, tangible objects are not relevant in digital environments and they can become a burden. A digital collection is - or it should be - free from the restrictions of tangible objects. Jason Donovan talks about this problem:

> In museums, the primacy of the authentic object is paramount. In cyberspace, it's moot. The object is not available except in surrogate form, so why organize content in an object-centric manner? People do not care so much if you have a certain painting, artifact or specimen as they are fascinated by all the context, the history, associated with it. In other words, what people want to know is not the What as much as the Who, Where, When, How and Why. Object oriented information is at too fine a level of granularity to be of use to most end-users. (Donovan, 1997)

Museums are nowadays information processing institutions specialised on (cultural) historical information. Exhibitions, tours, websites, online-databases, educational materials and different kind of community projects are all results of information processing. Although museums have acknowledged that they were dealing with information before, the radical change has happened in targeting and role of information. Information processing in museums was earlier for inner use only. It was meant for curators, conservators and researchers in order to help out exhibition design, collection management and research.

> A museum with a sizable collection of cuneiform tablets, for example, could provide online access to thousands of database records containing basic information such as measurements, date created, and location found, but such information, while valuable, would likely be of use to only a handful of scholars. Most visitors, especially students and children, will be more interested in learning about cuneiform writing in general, with few chosen artifacts selected as prime examples. Building such interpretive, educational information resources, however, can be far more difficult and time-consuming than simply providing access to database records. (Marty, 2008b)

When museums have started to open their collection data, the internal nature of that data has become evident. Collection management is not the same as cultural historical documentation and pure collection data lacks of cultural historical context.

If we analyse what kind of information we need to store in cultural heritage information systems, we will see that there are different kind of layers of information. The actual **management layer** involves "hard facts" of the object. The

situation where a museum documents its own functions produces information that is certain and which is not subjective. When did museum receive the item? How much does it weight? When and how was it restored? Where is it located? When was it loaned? This kind of information is produced by the museum itself and formulation of data can be shared with all the items in the collection. Relational databases are excellent for storing and handling this kind of information.

The second layer can be called a **property layer**. Each type of item has different kind of properties. For example, a building has a property that links it to a certain architectural design or a machine part is linked to the whole machine. The question is not about direct properties of tangible items, like weight, but more abstract properties. This leads back to the issues in data modelling. If there is no concept of architectural design in the system, then there is no way to connect the building with its design. All these properties have their own semantics that must be expressed somehow.

The third layer is the **conceptual layer**. This level connects items to to wider concepts and it involves interpretations and subjective views. These concepts can be keywords or semantic structures or some external documents, for example. We can say that a bicycle is a transportation device or that a person is a biological creature. These conceptual structures can be stored inside the system or they can be external. Defining these concepts is not a trivial task (see Conceptualisation and ontologies). In addition, certain terms might have local meanings and therefore it is important that there is a possibility to document these concepts also locally.

The last layer is the **contextual layer**. This layer is mostly referring outside from the system to our cultural conventions. For example, the title of a religious painting might refer to a certain section in the Bible. The context of the title is then the key for interpretations of that painting. It is not possible to document the meaning of the title in an objective way but it is possible to refer to the origins of that meaning.

I am not arguing that these layers are strictly separate things. I am using this separation as a tool to demonstrate different kind of information levels that we are confronting in the digital heritage. These levels can be found by careful ontological analysis. If all these four layers are combined in a same information system, the end result is a complex system that stretches the traditional database design to its extreme. The item-centric tradition in the museum documentation practises is the major road block to developing information systems in the museums. It prevents systems for including concepts and knowledge, or at least it makes the handling of those extremely difficult and inefficient.

Resources and output

Information processing needs resources. In the foreseeable future there are no any indicators that the funding of museums would receive any significant boost. The trend seems to be quite opposite. Currently Finnish museums are confronting cuts in their budgets and some museums in the USA have been even closed. Still,

museums are receiving new tasks and are confronting new demands. This means that there are less resources per task.

What is the output of the museum? By the term output I mean all activity that is intended for the outside world: exhibitions, tours, websites, online-databases, educational materials and different kind of community projects. Although the term is a little bit technical, it characterises the this side of the museum work: Output is that part of the organisation's work that is visible for common people. There are aspects in museums' output that can not be measured easily like how the museum affects in the society. But there are several others that can be studied and even measured like how many visitors the museum had, how exhibitions were received or how many used the museum's website and how. These measurement are used in many museums as an performance indicators.

The performance indicators have important role in opening digital heritage. If the website visitor count is one of the aspects that affect funding, then there is no financial motivation for the museum to open their materials for re-using. If materials are copied and used outside their website, there is no increase in visitor count of the museum website and hence there is no financial reward to the museum.

It is also possible to see the situation in a quite inversely. Instead of sharing their materials, museums could integrate information from separate sources. Diane Zorich writes:

> Museums offer collections, information, and context to their audiences. These are desirable assets on the Internet. But as Internet usage expands and more individuals and institutions become information providers, there is an increasing need to integrate the wealth of information that is available on-line. Museums could position themselves to do this for cultural heritage information and thereby carve out a useful niche on the network frontier. Unfortunately, in their initial forays into a networked environment, museums on-line sites have concentrated on presenting institution-specific information. This approach will not be tenable for long. Users of on-line information are demanding better (versus more) information, and Internet usage is moving from "browsing" to "searching". Accessible, integrated resources will be in demand as these trends continue. (Zorich, 1997, p.194)

Zorich is seeking museums that integrate information from different sources and offers that compilation to online users. The article was published 1997 and yet, as writing this in 2011, this has not happened. Museums are still providing institution-specific information that is not linked to other sources. One obvious reason for this is the lack of resources. Exhibitions are one of the most expensive enterprises in museums (McLean, 2004), designing websites and online materials requires - often costly - experts, and community projects usually requires a lot of human resources. Collecting materials from constantly changing the Internet and combing them with institution's own materials would need considerable efforts.

Stranger, guest, client or something else?

The outcome of museums has changed radically over the time and this change is reflected in a way museums have seen their visitors. According to Zahava Doering there are three attitudes that museums can take towards their visitors: stranger, guest or client.

> Strangers: This attitude arises when the museum maintains that its primary responsibility is to the collection and not to the public. Guests: From this point of view, the museum wants to "do good" for visitors primarily through "educational" activities. Clients: In this attitude the museum believes that its primary responsibility is to be accountable to the visitor. (Doering, 1999)

Doering states that transition of modes is not a sequential development through the history of museums although the modes have emerged during the history. All three modes still exists today and they can be found across institutions.

In stranger mode museum sees itself mainly as a research and preservation institution. Visitors have access based on "Doors are open" principle: While there may be visitors, there is no obligation for museum to react to the needs of visitors and visitors are not part of museum's core activities in any way. This mode can be found also in museums' digital assets. Like Kevin Donovan argues, when a museum "... plug the database into the Internet for all to see", that really does not provide access to a meaningful information to any other than a very specialised audience that knows what to search and how to do it (Donovan, 1997). By opening the doors of their internal database the museum does a very little favour for average users. Still, those raw materials could be the basis for some new modes of activity.

> Simply providing the public with access to data is insufficient to satisfy the goal of public education. (Donovan, 1997)

All three modes assume that museum visitor is definitely their visitor and that visitor is human. As long as we are talking about physical visits or using institution's website, that is a correct assumption. But what if we stop thinking about visitors? What if we have new quarters that do not want to visit but to use materials. What happens if we add a new mode. A mode that is interested of using museum's material in his/her own purposes? A mode that escapes framework of museums? These new participants could be called agents or data hackers. They are able to build new services from raw materials and they use materials where ever they can be found. These agents or data hackers are not limited to one source. With digital access one can built applications using several different sources.

Current online users are already escaping from these three roles by Doering. They are not necessarily interested of institution itself but only the content that institution holds. In this sense they are nor guests or clients. This is how Mike Ellis describes new users of the museums:

> We're only just starting to understand that our users are **talented** and **fickle**. They get what they want from whoever will provide it. They want **more** stuff via **more** channels.
>
> this means users **find** stuff through means other that those we provide, they **drop** into the depth of our sites rather than via the homepage and they **use** our material in ways we'd never even begin to imagine.
>
> and increasingly, they're experiencing **our** content somewhere other than on **our** site. (Ellis, 2009)(emphasised by the original author)

The whole question of the role of the visitor or the user, is related to the question of the role of the museums. Why do they exist and who should be able to define what museums do? If museums are seen as an instrument for public good, then it is possible to argument that this instrument should be as open as possible in order to fulfil that mission. Stephen Weil arguments that community that is funding museums should have something to say about the role of museums:

> The other attitude in need of change involves the museum's relationship to the community. The emerging public-service oriented museum must see itself not as a cause but as an instrument. In some considerable measure, the cost of maintaining that instrument is paid by the community: by direct community support, by the community's forbearance from collecting real estate, water, sewer, and other local taxes, by the considerable portion of every private tax-deductible contribution that constitutes an indirect public subsidy from the community. For that reason alone, it might be argued, the community is legitimately entitled to have some choice-not the only choice, but some choice-in determining just how that instrument is to be used. (Weil, 1999)

Tomislav Sola presents three directions that heritage communicators could take. These directions define not only the role of the museum but also a relationship between museum and their users.

> 1. The first direction (INTELLECTUS) continues the original one and will always be the easiest to choose. The perspective it gives is the perfection of scientific facts, perfection in specialization, constant growth in size and numbers and satisfaction by the existence of (our own) public which greatly esteems the knowledge and expertise concentrated in museums.
>
> 2. The second direction (EUPHORIA), which goes down the slope to what looks like a beautiful valley (therefore as easy to choose as the former one), leads to the leisure business and well-paid entertainment (if one beat the rising competition).
>
> 3. The third direction (SAPIENTIA) is not a road maintained and compacted by organized legions of travellers, but a rising path used by a few. It leads to hybrid institution compose of many and even divergent interests, serves a s place as a of exchange and getting together where teaching and learning, besides happening simultaneously, merge into understanding. Those who walk that path may shyly admit that their simple goal is to make the world better. (Sola, 1997)

The first direction by Sola can be interpreted as a traditional path, where museum is providing information to the public in the terms of the museum. The second

path describes the path of some kind of disneyfication, where heritage is coated with easily digestive coatings and where unpleasant materials are removed. The Sapientia path, in turn, aims for understanding. This path is not based on one truth or one museum. Sola is still talking about institutions and their role but the Sapientia path could be also see as a more inclusive path that also includes "outsiders" as contributors.

3.2 Conceptualisation and ontologies

Conceptualisation and formal ontologies are essential parts of information management in cultural heritage context. Cultural historical materials are difficult from the conceptualisation point of view. The very essential question is: How is a current target linked to other items, ideas, events and persons. And secondly, how those links and concepts can be formally presented so that they can be used for information storage and information retrieval purposes. Gruber defines conceptualisation process as follows:

> A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. (Gruber, 1993)

The idea of conceptualisation is not to capture all the aspects of a target but to find the essential ones that can be represented. This representation can be done with formal ontology. Clay Shirky presents the most common definitions of an ontology:

> It is a rich irony that the word "ontology", which has to do with making clear and explicit statements about entities in a particular domain, has so many conflicting definitions. I'll offer two general ones.
>
> The main thread of ontology in the philosophical sense is the study of entities and their relations. The question ontology asks is: What kinds of things exist or can exist in the world, and what manner of relations can those things have to each other? Ontology is less concerned with what is than with what is possible.
>
> The knowledge management and AI communities have a related definition – they've taken the word "ontology" and applied it more directly to their problem. The sense of ontology there is something like "an explicit specification of a conceptualization."
>
> The common thread between the two definitions is essence, "Is-ness." In a particular domain, what kinds of things can we say exist in that domain, and how can we say those things relate to each other? (Shirky, 2005)

The first definition provides a good bases for understanding ontologies in cultural heritage context: What kinds of things exist or can exist in the world, and what manner of relations can those things have to each other? With formal ontology we are able to present knowledge in a form that can be shared with any creature capable of logical reasoning. If we have an ontology stating that a person is a biological object (i.e. a person is a subclass of a biological object class), then

this is true regardless of the viewer's own beliefs. We have then shared conceptualisation of a person. However, this does not mean that we have captured the essence of a person. We are merely saying that every person is inevitable also a biological object.

Important notion is that ontology is about consistency, not about completeness. There can be knowledge that can not be represented with chosen ontology but all that can be represented is represented consistent way. In the digital heritage field, ontologies have been seen as a tool for information integration (Sinclair et al., 2006). When different sources uses same background ontologies, it is easier to combine those sources. But this has overshadowed the other side of ontologies namely that they can be used in order to improve the quality of the documentation. Ontological analysis is an intellectual process where the target is conceptualised and relations between terms and concepts are defined. Following examples of architectural drawings and Gothic style illustrates this process.

> For example, in the case of architectural drawings, it can be said that the immaterial architectural design is carried out by physical drawings. This way the actual design can be documented independently from the physical documents. This separation also helps when modelling the relation between industrially made objects and conceptual designs. Industrially made furniture is produced by following a certain procedure or design. The design is documented only once even if there are several physical copies of the furniture. (Häyrinen, 2011)

An architectural design is carried by architectural drawings and a building is produced based on that design. The design is created by an architect, drawings are drawn by someone (could be the architect or some one else) and the building is constructed by construction company based on that design. The architectural design could be based on some earlier works of the same architect or the design can be part of an architectural competition, for example. Careful ontological analysis is needed in order to reveal all these relations.

In order to further demonstrate the conceptualising issues, I will present the following example of Gothic style. Now, what do we mean by Gothic style? Is it a style or is it a time period or is it both? If Gothic style is modelled as a style, then there are some common characteristics that defines the Gothic style. If we apply those characteristics to an architectural design today, do we have a Gothic design? No, since Gothic is also a time period and therefore it is not possible to build Gothic architecture never again. But if Gothic style is a time period, does that mean that all building of that time period are Gothic? Again, no, since the Gothic style is not only a time period. What about location? There were no Gothic style outside Europe, so that adds location to the equation of Gothic style.

If the term "Gothic" is limited only for buildings with certain characteristics that were built during Medieval period in Europe, it excludes the use of the term in other contexts. Gothic features can be found also elsewhere and the term can be used for describing architecture that emphasises verticality and light without any relevance to the Gothic period itself. In other words, the term "Gothic" is more useful if it is not nailed down to a certain historical period but instead it is

used to define some generic properties in spite of the location or time period. It is different to say that a building has Gothic features than to say that the building is a Gothic building.

So, what if we want to find “real” Gothic architecture? What defines real Gothic buildings is the combination of construction date, location of the building and the building’s stylistic characteristics: Real Gothic buildings are those which have Gothic style characteristics, are located in Europe and were constructed somewhere between 12th and 16th century. The concept of the Gothic Architecture is now in a form of query and it - at least in a very loose, information retrieval sense - defines the term Gothic.

Ontological analysis is not an easy task. We have different meanings for things even if we think that we have shared understanding of a topic. Clay Shirky presents excellent notes about the problems of finding shared meanings:

> The mind-reading aspect shows up in conversations about controlled vocabularies. Whenever users are allowed to label or tag things, someone always says "Hey, I know! Let’s make a thesaurus, so that if you tag something ’Mac’ and I tag it ’Apple’ and somebody else tags it ’OSX’, we all end up looking at the same thing!" They point to the signal loss from the fact that users, although they use these three different labels, are talking about the same thing.
>
> The assumption is that we both can and should read people’s minds, that we can understand what they meant when they used a particular label, and, understanding that, we can start to restrict those labels, or at least map them easily onto one another.
>
> This looks relatively simple with the Apple/Mac/OSX example, but when we start to expand to other groups of related words, like movies, film, and cinema, the case for the thesaurus becomes much less clear. I learned this from Brad Fitzpatrick’s design for LiveJournal, which allows user to list their own interests. LiveJournal makes absolutely no attempt to enforce solidarity or a thesaurus or a minimal set of terms, no check-box, no drop-box, just free-text typing. Some people say they’re interested in movies. Some people say they’re interested in film. Some people say they’re interested in cinema.
>
> The cataloguers first reaction to that is, "Oh my god, that means you won’t be introducing the movies people to the cinema people!" To which the obvious answer is "Good. The movie people don’t want to hang out with the cinema people." Those terms actually encode different things, and the assertion that restricting vocabularies improves signal assumes that that there’s no signal in the difference itself, and no value in protecting the user from too many matches.(Shirky, 2005)

The point that Shirky makes is that although we might think that certain terms can have the same meaning for everyone, this is not true. The terms may have closely relations but they mean different things. Shirky’s examples are revealing. Even the first example of his shows the problem. The Apple is a company, Mac is a generic name for Apple’s desktop computers and OSX is an operating system running computers made by Apple. So all these three terms mean totally different things although they are related. In certain level these terms could be

used as synonyms, for example for tagging purposes. The problem is that for certain people these terms are synonyms and for others they are not. This is the problem of global ontologies. Terms might have different, local meanings and if global ontologies are used instead of those local meanings, the information is skewed. What is needed is that also concepts used by the heritage information systems should be documentable units. These would provide a way to define the actual meaning of the term from the user's point of view. However, this does not prevent connections from local concepts to global ontologies.

3.3 Long-term preservation

Digital heritage may be the most fragile heritage that we have. A crack in the ancient vase does not destroy the evidential value of the vase. The vase might be more fragile than before but as a whole, it still has all the heritage characteristics left. But a "digital crack" - missing sequence of bits - in digital data may render a digital object totally inaccessible. Long-term preservation is described by Janee, Mathena and Frew as follows:

> Long-term preservation requires solving two complementary, yet coordinated problems: preservation of the information itself (the raw "bits"), and preservation of the context surrounding the information (the meta-information needed to interpret the bits). (Janée et al., 2008)

Like the definition of digital heritage by UNCESCO, also above definition of long-term preservation is simple and easy to understand. First, raw bits must be preserved. Binary data must be kept intact since even a change of one bit might render whole file unreadable or prevent application to be executed. Secondly we must preserve the language of the bits (data format) that tells us how bits need to be interpreted. These two tasks form a very core of the long-term preservation of digital heritage.

In addition to these purely technical aspects of long-term preservation, there are several non-technical issues. We must answer questions like who is responsible of long-term preservation and what should we save. The rapid and unpredictable growth of technology, the changing economic and political climate, and the legal and rights management regimes that control digital content (Lunt et al., 2008) are factors that also affect to practical long-term preservation. Despite the open questions and obstacles, long-term preservation needs to be done.

> We are at a unique point in history, where cultural heritage professionals must work to care for the physical past while assuring that there will be a digital record for the future. (Mudge et al., 2007)

In the first phase we must ensure that we are able to preserve the raw bits without errors. This can be done with live digital environments that are kept running with careful backups or bits can be transferred to a physical medium that is stored for

long-term preservation similarly of tangible heritage objects. In practise, materials should be accessible so they must be located on electronic environment even if they have been "burnt in" some non-electrical medium.

There is an obvious difference between digital documents and computer applications when it comes to long-term preservation. A document - like image, video or text document - holds information. If the file format of a document is documented, it is possible to access the information that document holds. Therefore open and well documented file formats are crucial for long-term preservation. Only when a reasonable length documentation of the file format is freely available, it can be guaranteed that files could be read in future. But for applications the preservation task is more complex. Applications includes machine interpretable commands and possibly human-computer interaction. There is no open file format for applications and even if there were such format, it would not solve the problem since applications are linked to platforms and their mechanisms. The executable file of the computer game is preserved as a game only if we have ability to actually play it. Otherwise we have preserved "a silent item", like a book that was written in a language that no one can not understand anymore.

There are three main approaches for digital preservation: technology preservation, migration or emulation. Technology preservation simply means that we preserve hardware in order to run applications on them. Since electronic devices have a certain time-span, this is not a viable solution for long-term preservation. Hardware preservation requires knowledgeable maintenance staff and fragile machines can not be lend the hands of a wide audience. However, a centralised hardware preservation can offer possibility to open exotic file formats that can not be opened other ways. A good example of this was the spare part catalogues of Finnish Hornet fighters. Originally, these catalogues were in punch cards. These were transformed to a magnetic form by the Finnish Data Processing Museum Association with a punch card reader that was in their collection of old computers. (Pitkänen, 2007)

Migration means that the digital item is replicated with other technique maintaining the original appearance and functionality as strictly as possible. Images can be re-saved in other format or video can be compressed with a different compression algorithm. Old computer games have been migrated to other platforms, like the browser version of the PC-computer game Doom II.

However, migration of computer applications is problematic from the cultural preservation point of view. Obviously, a constructed replica is not an original work and then we might ask what was preserved. It is like a copy of statue or painting; it looks like the original but we know it is not a real thing. Besides, replicating of any interactive software is not a trivial task. Basically the whole software must be re-written, the user interface elements must be re-made and, in the case of games, the logic of the gameplay must be discovered and then re-implemented. If the source code of the original software is available, the task is little more easier. But if there is no access to source code, the task is usually too time consuming for actual preservation purposes.

In the third approach, called emulation, an application, an operating system or a piece of hardware is emulated (in the last case I prefer the term "virtualisation" since hardware is turned to be virtual). A good example of emulation approach is a Commadore emulator called Vice. Commadore-64 was introduced in 1982 and it become popular game platform with several million devices sold. It can be said that there is Commadore generation, people that have spent part of they youth with gaming and programming with it. Now, after 30 years there is definitely nostalgia attached to Commadore-64. Vice allows gamers to play those games again. The purpose of the emulation is to replicate the original behaviour. A good example can be found again in VICE and its file loading. Loading files in current computers is much faster that it was in C-64 which used floppies or tapes. Instead of loading files instantly, VICE has an option to replicate the original speed of floppy drive thus offering as original experience as possible with appropriate loading pauses for retro gamers (see also C64.com in Cases chapter). But it also good to point out that emulation might not be able to capture all the details of the original work. For example, games that were designed for TV screens look different on LCD monitors.

> In today's world of huge, sharp LCD monitors, it's hard to remember what a videogame image looked like on an ordinary television of the late 1970s. Emulators like Stella make it possible to play Atari games on modern computers, serving the function of archival tool, development platform, and player for these original games. But unfortunately, they also give an inaccurate impression of what Atari games looked like on a television. (Bogost, 2009)

Emulation can be used to help to preserving interactive digital items. However, it is not a route that can be guaranteed to work. There is very little commercial need for emulators and actually most of the emulators are made by volunteers. The target system may be poorly documented and emulation must be built with trial and error. For those reasons, the emulation maybe not fully implemented. It may be able to run certain applications but not others. The worst case would be a emulator that renders emulated application in a way that radically differs from the original experience. Then, of course, the whole process of preservation is rendered invalid.

Virtualisation goes beyond emulation by emulating hardware. Computers are based on different processor architectures and software that was compiled for a certain processor architecture will not work in other architectures. In virtualisation, a virtual hardware is created to the top of a physical hardware with a virtualisation software. Then an operating system is installed on the top of virtual hardware and finally applications can be installed on the top of the operating system. This means that applications run in their native environment and old application can be still used. However, the ecosystem needed in the virtualisation based preserving is complex. We need virtualisation software, operating system and its application libraries (and right versions of them), the original application and its supporting libraries.

From the point of view of museums, long-term preservation is a challenging question. All the technical issues related to long-term preservation make it hard

to understand without technical background. Still, the problem is even more current for small organisations. There is a immediate risk that all museum's digital materials are lost because of hardware failure or corrupted CD-roms, for example. The first step of long-term preservation is naturally making backups of the data. If that fails there might be no materials to preserve.

But if we accept that there are relevant and important cultural historical information sources outside official heritage organisations, then we must also re-evaluate the long-term preservation practises in these organisations. If museums only take care of preservation of their own materials and exclude external information sources, then we are in danger to lose the richness of those external sources. This means that the second democratization process of heritage disappears from official archives. It is - again - the official and exclusive truth that gets stored for future generations.

3.4 Rights

It is almost impossible to write about digital heritage without any mention of copyright. In the museum field a persistent problem is that copyright owner is not known. A part of the the problem is that the donor is not necessarily a copyright owner and thus cannot grant any rights over the original copyright owner. Copyright problems are especially related to online access of digital surrogates. Copyright sets restrictions about what can be published or when that can happen. These issues are complex and as Akmon states, there are different kind of approaches that organisations could take:

> Archival institutions employ several different strategies to respond to the barrier that copyright law represents to online, open access: they avoid digitizing collections with complicated rights issues (e.g. not in the public domain, copyright not held by the archives, or copyright held by a third party); they interpret their digitization and distribution of materials online as ''fair use'', or they attempt to identify, locate, and seek permission for every item they plan to put online. (Akmon, 2010)

These approaches are: digitise all, digitise only when copyright is owned by archive or digitise only when permission is granted. I should note here that by the term digitise we mean also a possibility of online distribution. Merely digitising for internal archive use is rarely problematic but the problems start with online accessibility of the material.

Although Google is not a a museum, it has an interesting case of digitise all approach: Google Books. Google Books is an example of hacker-style problem solving. Sergey Brin, the co-founder of Google with Larry Page, said that "If we had done that [ask permission] we might not have done the project" (Auletta, 2010, p.96) Permissions were not asked which lead problems later with book publishers. The selection what is distributed online is made according the copyright: All books are searchable but the full content is available only for books that are out-of-copyright while copyrighted books are available only partially.

Studies show that tracking copyright and asking permissions to distribute items online is resource hungry task. Depending largely on selected method for requesting grants and the type of the material, the cost can be considerable high per title. And even then, there is no promise of solution. In the case of Cohen Collection with 5463 copyrighted items, of all of the responses from rights holders, 91% were to grant the Library permission to display all items requested in permissions letters. Still, only 64% of the copyrighted items could be displayed online. The reason was that even though the copyright holder was found, there was no guarantee that there were any response to the request. In 32 percent of cases there was no response at all. As a result, 18 percent of the collection could not be shown online because of the non-response of copyright owners and 13 percent of collection could not be shown because a copyright owner was not found or copyright information was not found. When summed, this means that 31 percent of the collection is inaccessible, not because of the copyright holder's denial, but because of inaccessibility of copyright owner or copyright information.

> The result of this type of response is that the Library ultimately had permission to display online 3,490 (64%) of the copyrighted items in the collection. Unfortunately, 981 (18%) of the copyright items in the collection could not be displayed due to non-response from rights holders. Another 687 (13%) could not be displayed for three main reasons: staff could not identify the rights holder (22); staff could not locate the rights holder (309); or the rights holder was a company that they found to be defunct (356). Only 294 (5%) of the copyrighted items in the Cohen Collection were explicitly denied. (Akmon, 2010)

In the case of Carnegie Mellon University Libraries, a random sample of 277 titles from the collection was selected copyright inspection. Twenty one (21) percent of the publishers of the books could not be located and more than a third (36 percent) of found publishers did not respond requests at all (Covey, 2005). As a result, only about a fourth of the sample titles could be published online.

The idea of fair use is commonly adapted by Internet communities. The fair use is an exception in a copyright law that allows the use of copyrighted work without the permission from the copyright holder in certain cases. The legal status of fair use differs by country. For example US legislation recognises fair use(U.S Copyright Office, 2009) but, for example, the Finnish legislation does not. There is right to cite copyrighted works in Finnish law which reminds in principle the idea of fair use but is more restrictive.

The adoption of fair use among the Internet users means that materials are re-used no matter of their copyright status. Maybe the most familiar example of this kind of re-use is Youtube. It holds a lots of copyrighted materials without permissions from copyright holders. The lack of the doctrine of fair use in the European copyright legislation makes most of these materials illegal. Currently Netherlands is considering changing its copyright laws to include the fair use doctrine. Bernt Hugenholtz of the Dutch state committee comments about Youtube regarding the current copyright laws:

> "We all agree that it's good for creativity, good for laughs, and no one gets hurt. Copyright holders are not harmed, so it makes a lot of sense to allow

> this. But in Europe, where we do not have open norms like the fair use doctrine in the United States, we can't do these things without infringing the law." (Chesal, 2012)

It is good to remember that there are huge amount of materials in museums that are no longer covered by any copyright. Making these materials available online is then a question of resources and the attitude of the collection owner.

3.5 Information integration and semantic web

Information integration in this context means methods for harmonising different kind of cultural historical information sources so that they can be used together. Harmonisation can be accomplished by using shared data models between databases or using different kind of metadata formats. Lately, ontology-based approaches have been emerged like CIDOC-CRM.

It might be surprising that information integration is introduced in the problems section of open digital heritage. After all, the purpose of information integration is to enrich materials by combining them and creating unified search and browsing facilities. However, from cultural heritage documentation point of view information integration is both an opportunity and a risk. The opportunity is that materials and information from different sources can be collected, combined and searched more easily. But the risk is that information integration is seen as an intrinsic value that affects documentation practices and tools by limiting the flexibility of documentation. Instead of creating tools for cultural heritage documentation and taking account local needs, openness and practices of heritage institutions, this can lead to a situation where tools are created solely from information integration point of view.

There seems to be some kind of "more is better" ideology behind the information integration efforts in digital heritage. In another words: more information in same place is also better information. It is true that more information is better information for example while searching artwork of a certain painter: The more paintings are found, the better the result is. Another point of view is that all those millions records that are not in the area of interest for the user, are extra load that can "hide" those things that the user is searching for.

This is one of the main questions of information retrieval: How to find relevant records among millions on irrelevant records? A blank search field might not offer the best user experience for a user who wants to find information about certain topic among millions of records. The problem is that user does not know what terms should be used (or other way around: the user does not know what terms the information producers have used).

Semantic techniques are introduced in order to improve the situation. So far, the semantic web has been a promise with very little practical success stories. This may change in the future, but there are some serious hurdles in a way. Creating semantically aware content is not a trivial task. In many cases it is a labour

intensive task and concepts and paradigms are not familiar for example average museum workers. Building ontologies and them mapping existing materials and databases to them is a complex task.

The mantra for semantic web is that it is able to contextualise things. With machine readable semantics, machines are able to create rich content for digital objects. At its best, this can create browsable, rich and even surprising digital heritage resources. In worst case scenario, this can lead a situation where information overload is replaced by context overload. Context that is not relevant to the user is overload. For example, when viewing an artefact made of wood, then do I really want see all the other objects made of wood? The problem is that this kind of context can be easily generated automatically. For instance, every mention of the city of Paris in a record can be linked to the map of Paris, to the home page of Paris city, to the Wikipedia page of Paris, to the images of Paris in Flickr and so on. It is easy to drown the user to the endless sea of automatically created context. Still, it is really difficult to create the context according the user's needs. But this is something that other persons with similar backgrounds and interests could create. This is possible only if those people have opportunity use material freely in order to build their own sites, blogs and virtual collections.

3.6 Summary

It is not a coincidence that I placed museums in the chapter called "Problems of Open Digital Heritage". Museums are part of the problem for several reasons. Their information management is mainly internal, item-centric and meant for experts, there are insufficient understanding of possibilities of participatory web, there are fears for misuse and fears of loosing authority and there are problems with copyright. Museums can not be blamed all of these problems but some of them - like documentation practises - are "museum-based" problems. Still, according the museum literature museums seem to understand quite well the sore spots of their practises when it comes to opening digital materials. Museums see the change in public expectations, the problems of item-centric documentation are admitted and resource problems are almost self-evident. However, there seem to be two aspects that are not covered.

In museum 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. Even in projects where people can participate - like crowdsourcing projects - people can participate only by the terms set by museum. The idea that museums make all the work - and should make all the work - is based on the idea of an expertise organisation. Because the museum is an expert when it comes to their materials including digital data, only the museum is allowed to manage that data. A problem is what happens when museum has no resources to do it? Nothing. If museum is not able to provide context to their data then no one else should do that neither. In practise this is the current situation in many small

museums: museums can not contextualise their collection data because there is no resources. Others are not allowed to do it since they are not “insiders”.

The second very basic assumption - that is related to the first one - in museum literature is the existence of an client or visitor. Although there are discussions about the changing role of the client, there are no doubts that client exists. Museums seem to be fixed to the idea of a client that uses their resources in the way that is defined by the museum. This makes sense when there is a traditional or virtual exhibition by the museums. But if materials are copied and used outside the organisations - which is happening when materials are online - , then there is no longer visitor or client but only anonymous person who is consuming, remixing, combining or sharing materials. This use of materials falls out of the context of museums discussions.

However, recently also non-visitor based approach has taken shape in practise. Several museums have released their materials under free licences, for example in Flickr Commons. When museum releases their materials under free licence, then the museum has given up of thinking itself as an organisation with clients, but as an organisation contributing to the large pool of open digital heritage. This is a very important change. The obvious problem of this schema is that the museum do not necessarily get credit for their work. Images, for example, provided by the museum can be used in some website or blog without crediting the provider of the images. But the credit of the museum is that they have enabled something that would not have been possible otherwise. Museum can be credited for providing starting points for new interpretations and new uses of heritage materials.

In the end, all depends on how museums see themselves and how they see the Internet. If museums think that the Internet is just a place for a website of the museum where they can promote their exhibitions, then there is not much that can be done. The Internet is then just one channel more for the museum’s media portfolio. But if the Internet is seen as a new way of doing things, a new way to share and produce information and a new way to bring cultural heritage to more close to people, then open digital heritage would be possible.

4 CASES

I have chosen cases that exemplifies of what is currently happening in the digital heritage field including non-institutional quarters. I have selected an offline example (Tipitii exhibition), a case that raises strong feelings (Anne Frank museum), a case that demonstrates how technical implementations affected the accessibility of the data (Finnish national suites), a case that shows how museums are bypassed as an online information sources (Goya) and a case that exemplifies the structural disadvantages that museums have (C64.com). In addition I have a case of computer program called Blender as an example of open source community. Together these examples illustrates what is happening in the peripheral of official digital heritage: Outside of museum websites, outside of databases, outside of institutions.

Part of the cases are in the separate article called "Google, the world's largest museum?" in the conference papers section of this book.

4.1 Blender

Blender (or Blender3D) is an open source 3D-modelling, animation, game development and video-editing software. The Blender community is large enough to form an ecosystem. When Blender gains popularity, the more there is a need for training materials which opens route for commercial services. And more there are trained users, the more attractive Blender is for commercial studios.

The Blender community can be divided to two sections: the developer community (kernel space) and the user community (user space). The difference between these two groups lies in the nature of output, accessibility and in the number of participants. Developers work with the final product and they have to have an agreement on what is done and how. In other words, there can be only one output from the developer community which is the final program. Therefore there must be some kind of controlling system for monitoring and controlling the participants of the kernel space. In the user space, however, this kind of control

is not needed. Anyone can participate, for example by creating a tutorial or by writing a book.

Blender's developer community is not large, which is the case in the most open source projects. Inside the developer community responsibilities are divided according to each developers activities. A developer can be maintainer of certain section of the code and s/he can quite independently make decisions about that section. However, since all developers are working with same source code, there are limitations of what individual developer can do without breaking the whole system.

However, the user community of Blender is quite large. There were more than 14 millions visits in the blender.org from April 2010 to April 2011. In 2009 there were 3.4 million downloads of the software. Since it is a free program and thus anyone can redistribute, the actual download numbers are much higher. These figures shows only downloads from blender.org. There are several discussion forums related to Blender in different languages. The largest is Blender Artists forum which has over 85 thousand registered users. Hundreds of Blender tutorials and screencasts can be found with Google search. In addition, there are several books available and also commercial support and services for Blender.

The relationship between developer and user communities is interesting. Users have very different backgrounds and needs which leads to a great variety of demands from the users. Some users want new features, some complain user interface issues and some can be very reluctant to any changes. Open source communities are not democratic systems. Users can raise issues and they can organise voting but the final decisions are made by the developers. Since most of the developers are working in their spare time, there is no obligation to please the users. The developers are an elite group with the power to define the future course of the software.

I discussed earlier about the open source movement and its practises in cooperation. One of the major reasons for the success of the open source movement is that it allows very different kind of people to participate in their own terms. Here is an example list of things one can do with Blender:

- - use software non-commercially and commercially
- - distribute software
- - write tutorials
- - translate tutorials
- - translate software
- - help people in discussion forums
- - get help in discussion forums
- - create your own discussion forum
- - write a book
- - offer commercial services
- - write scripts (applications that can be run inside Blender)
- - report bugs
- - confirm bugs

- - sell software
- - write patches in order to fix bugs
- - make your own version of Blender (forking)
- - develop software inside community

Only the last item in that list (develop software inside community) requires permission or acceptance from the developer community. Before this, one must have proved that one is able to work with Blender source code. If we exclude last three items from the list, we will see that the remaining items are non-technical tasks that anyone can start doing without any formalities or permissions. These activities all benefit the project directly of indirectly. The results of different activities are harvested by search engines and they are presented in blogs, sites and discussion forums. Even though the community is very loose and individuals might not even recognise that they are part of the community, it is still a community effort.

From the digital heritage point of view this model of co-operation is interesting since it is not based on a strictly defined task, like for example most of the crowdsourcing projects and there is no heavy central organisation. It offers a model for inviting co-operation of digital heritage where users are free to determine their tasks and goals.

4.2 C64.com

C64.com is website, that is concentrated on old Commodore games. It provides an archive of Commodore 64 and VIC-20 games. Games can be downloaded and played via emulators. In addition the site has some articles and interviews of game programmers, music makers and other related people. The F.A.Q of the site states that "C64.COM exists purely to preserve the culture around the Commodore 64 that might otherwise cease to exist because the mainstream computer culture has moved away from the Commodore 64 in the early 1990s." And later: "Since C64.COM is a museum, we are archaeologists who dig vast digital (and concrete) archives and arrange our findings in the museum."

Legally c64.com is in a grey area. Games are still copyrighted by the companies that created them but files are still offered for download. The term "abandonware" is often used about this kind of computer programs that are still copyrighted but that are too old to be used in modern computers. The site says:

> "It is [copyrighted], but most companies don't mind their old products being spread as long as you don't make money with them."

There is an interested twist caused by copy protection of the games. Most of the games are cracked. Cracks are used to remove copyright protection from the game files and they can contain also game cheats or other modifications to games. One of the articles in c64.com defines cracking like this:

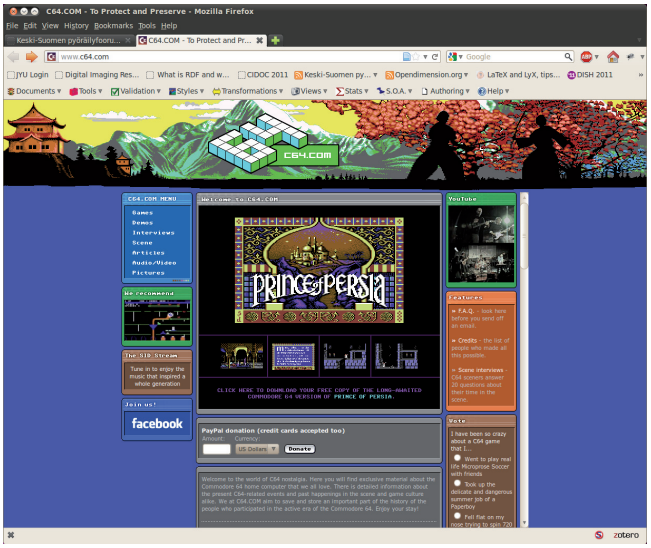


FIGURE 3 "To Protect and Preserve". The main page of the c64.com.

> Games are cracked, given new loading screens, boasting the name of the crew, and frequently compacted so they load quicker. On occasion the games are even polished up, improved in a way clearly beyond the original programmer.

This means that games in the c64.com are not original versions but they include modifications of the gaming community (or cracker community to be more precise). Therefore games have lost something of their original appearance. But in the other hand, cracked games are documents of the whole gaming culture of the Commadore games.

There seems to be no explicit long term preservation plan in the c64.com but at least a middle-range term preservation is done by sharing. As long as people are interested about these games and downloads and shares them, these games will be safe. Even if the whole site was deleted, it would be possible to collect most of the games back from the users of the site. There is no doubt that these people would be willing to help their free game museum.

The c64.com is the best resource for Commadore C64 and VIC-20 games. The site offers way to experience the games as they are best experienced: by playing them. For those who have played these games in their youth, the site is full of nostalgia. Those that are new to this game genre, the site offers a wide view to the gaming culture of 80's. C64.com demonstrates how non-institutional digital heritage sources have some structural benefits compared to institutional heritage organisations. Due the copyright restrictions any official heritage institution could not provide these games in playable form. The site is concentrating things that are relevant to the gaming community. There is no catalogue records of who "donated" which game and when that happened. Nor is there textual descriptions of cover art. But there are games that can be played. The C64.com has managed to create a community and they have reached target audience. That

is something that heritage organisations might want to take a lesson. As Melissa Terras puts this:

> 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 practise in constructing online resources, and reaching relevant audiences in the process. (Terras, 2010)

4.3 Anne Frank's House

Anne Frank's house - where she and her family were hiding from Nazis during World War II - is located in Amsterdam and is now serving as a museum. There is a virtual tour available from the museum's website that allows users to visit Anne Frank's house virtually. One can navigate through 3D-modelled apartment and see panoramic view of rooms. These panoramic views, that are constructed from photographs, are able to present inner parts of the house in a very detailed manner. The virtual tour of Anne Frank's house is an informative representation of the "real thing". It is able to give more information than just visiting the rooms themselves. However, it can not offer the feeling of actually being there, the presence of other visitors, noises and odours of the old house.

Even if it is possible to virtually visit Anne Frank's house - and even see the hidden parts - , it does not affect the originality of the real place. Quite contradictory, reproductions seems to work as an invitation. Everyone is - in the limits of digital divide - able to visit virtually Anne Frank's house but actual visit is something "more". Originality is not about information, originality is an experience. I can find a lot of information about Anne Frank, her history, her family and the place she lived during Second World War without ever leaving my apartment. But only original experience about Anne Frank's house can be provided by Anne Frank's house itself. Authenticity can not be digitised. It is a psychological aspect of the object.

However, from this study's point of view, the most interesting part of the Anne Frank Museum seems to be located outside the museum. The Anne Frank's museum has their own, moderated YouTube channel. The channel displays different kind of documentaries and interviews related to the destiny of Anne Frank and her family. Users are able to comment videos but comments are moderated. Comments in this environment are positive and encouraging. There are no offending comments, no racist or antisemitic statements or mocking of Anne Frank. But if we look outside the museum context, we can find similar videos also in the free Internet in unmoderated contexts.

The video called "Dear Kitty" Remembering Anne Frank Part 1 of 6 was uploaded to the YouTube by "lol2k20" in 30.05.2008. This video forms an interesting contrast with the official YouTube channel of the Anne Frank museum. Firstly, the

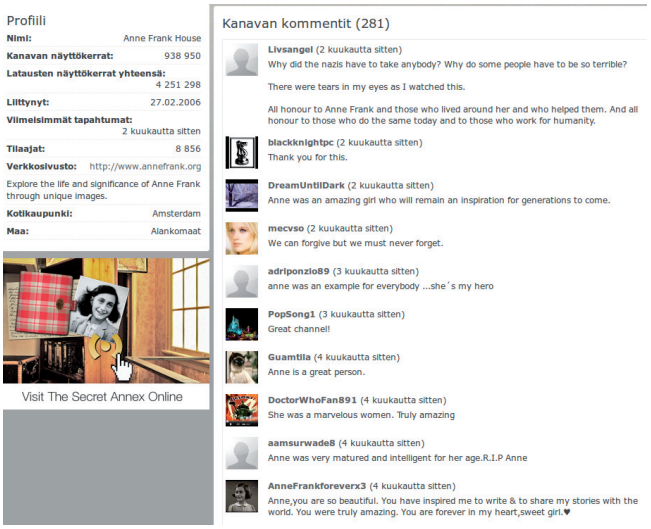


FIGURE 4 Comments in the official Youtube channel of Anne Frank Museum

video shows how copyright issues are seen in Internet and especially in Youtube. The person who uploaded the video, states in the video description as follows.

> Documentary in which Miep Gies talks about Anne Frank. *sorry about the quality but if i put it at a too high quality it would have taken ages to upload and for you to watch it* Also I do not own this documentary, If the owner is reading this, I will remove the videos.

Although it is recognised that the video is copyrighted, the user thinks that this video deserves to be viewed. Therefore “lol2k20” digitised it and uploaded it. As writing this (autumn 2011) the video has been viewed over one hundred thousand times. Comments section of the video is unmoderated. Most of the comments are positive and encouraging but there are also quite opposite comments. Here are few examples of the most radical comments of the video "Dear Kitty" Remembering Anne Frank Part 1 of 6:

> germany 1925 v usa 2010 germ usa check check - media run by lying heebs check check - jew financial scams lead to recession check check - jews run huge organized crime network check 2b done - nation wide program to stop this
>
> i hope the SS men got a chance to enjoy sex w anne before she died of AIDS. BlacksShdBSlaves
>
> – http://www.youtube.com/watch?v=kmAa2eN7wFY

This is the wild side of the Internet, no control, no rules. By looking the comments above it easy to see why unrestricted user participation can be problematic. When anyone can post anything, then anything will be posted. Most of the Youtube users have witnessed irrelevant and inappropriate comments by the other users. There is a term “Youtube troll” that means a person that leaves that kind of comments. The anonymity of the Web makes this easy. This anonymity

has two sides: anonymity also makes it easier for people to express themselves more openly. But, they can be also more cruel that they would in face to face communication.

However, there are some revealing nicknames in comments like "AbeLincolnStolMyNgrs" and "BlacksShdBSlaves". These clearly show that these user names are chosen to reveal racist motivations. Within the framework of this study it was not possible to study the true motivations behind these nicknames. But I think it is safe to assume that setting up the profile, selecting a racist user name and making racists comments is not just a hobby of an non-racist person but there must be at least some kind of racist motivation involved. If we agree with this assumption, then we must agree that these comments present a view where other humans are on a higher level than the others. And because of this ideology - more or less permanent - these comments are not just comments by trolls. Although these comments can be disturbing, they are not irrelevant. They are relevant because of the motivation of the writers.

After this short analysis of the comments in the everyday Internet, we can re-examine the Anne Frank Museum's Youtube channel. First I would like to point out that it is very understandable that the museum filters the comments in their channel. But we can also reflect the comments to the task of the Anne Frank Museum. What kind of world is presented in the Anne Frank Museum's Youtube channel? It is a safe place. It is a disneyficated reality with shared values and respect among people. It is a place where everything evil happened in the past. It is like a 3D-presentation of the museum itself, not a "real thing". But is that a problem? Why should visitors read comments that upsets them or make them angry?

I am arguing that one must see a value of free speech and free Internet even though it might expose for "bad things". If museum makes all the thinking and decides what is appropriate, then the role of the visitor remains passive viewer. The ability to make one's own judgements is not needed when everything is cleared out. I am not criticising the Anne Frank Museum's decision to filter the discussion in their Youtube channel. I am merely noting that, in a way, the museum's Youtube channel is a paradox. The museum is a memorial of persecuted people and one of the methods of persecution is the lack of free speech. Yet, this same right is missing from the museum's own media. The Anne Frank museum is filtering out that what we should remember. It is the wild site of the Internet that reminds us that the ideology that made Holocaust possible has not disappeared. We must be alert and best reminder of that is free speech.

4.4 Tipitii exhibition

I wanted to include offline example to my cases, which in this case was a traditional museum exhibition. The question here was if there could be such thing such as an offline exhibition. If we exclude to act of actually visiting the exhi-

bition, then what remains during and after the exhibition is the Internet context of the exhibition. This includes the museum's own materials that are "left" to the Internet but also materials that visitors create and publish, and also materials that are somehow related to the topic of the exhibition. In other words, all that material that is found when searching by the topic of the exhibition or the title of the exhibition, forms the Internet context of the exhibition. Mostly this context is out of the control of the exhibition creator.

Tipitii - Suomen broilerituotanto 50 vuotta exhibition (Finnish broiler production 50 years) was exhibited in Sarka-museum in 16.10.2009–31.10.2010. The exhibition was based on "Suomen broilerituotannon historia" written by Hilja Toivio. Tipitii-exhibition was sponsored by a commercial interest group called Finnish Broiler Association and this was clearly stated by the museum. While visiting in the exhibition, the overall impression was the exhibition was a quite typical museum exhibition with posters and some old and newer machinery. Broiler production was presented as an unproblematic part of ordinary food production. The title of the exhibition "Tipitii" and the studio photograph of baby chicken of the exhibition were messaging warmth and harmlessness. I think it is fair to say that attitude of the exhibition was uncritical and positive towards broiler production. This was not surprising if one looks of the sponsors of the exhibition. In the exhibition room there was a sign with the following text: "The Finnish broiler Association wish to thank the following: Suomen Siipikarjasäätiö ry, Atria Oyj, HK Ruokatalo Oy, Saarioinen Oy, Suomen Rehu/Hankkija Maatalous Oy, Rehuraisio Oy, Suomen Broileri Oy." All are companies - except Suomen Siipikarjasäätiö (Finnish Poultry Foundation) - that are operating on broiler industry.

In the Tipitii exhibition broiler production was presented as an unproblematic part of food industry with the exiting story of smuggling of the first broilers. However, in this study I am not deep analysing the exhibition itself. I am interested the difference between the perspective of the exhibition compared to what could be found online related to the topic and the title of the exhibition. I am arguing that the exhibition had - and still has - an online context although the exhibition itself was not online exhibition and that this context is out of the museum's control. Perspectives that were missing from the exhibition can be found online parallel to exhibition's online presence.

Broiler production is a controversial issue. Welfare of animals or animal rights issues in general are related to large production units and the breeding of fast growing broilers strains and health issues caused by that. The EU report "The Welfare of Chickens Kept for Meat Production (Broilers)" states:

> Most of the welfare issues that relate specifically to commercial broiler production are a direct consequence of genetic selection for faster and more efficient production of chicken meat, and associated changes in biology and behaviour. [...]
>
> A wide range of metabolic and behavioural traits in broilers has been changed by selection practises. Major concerns for animal welfare are the metabolic disorders resulting in leg problems, ascites and sudden death syndrome and other health problems. (European Comission, 2000)

Päivänikäinen broileriuntuvikko. Kuva: Annele Mäkijärvi

Broileri ei ole tavallinen kana. Se on lihantuotantoon jalostettu nopeakasvuinen kananpoika. Suomessa broilleria tuotetaan elintarviketeollisuuden ja maatalouden yhteistyönä.

Broilerituotannossa on monta vaihetta. Tuotantoketju alkaa broileri-isovanhempaisuntuvikkojen lennättämisestä Skotlannista Suomeen. Untuvikot kasvavat aikuisiksi linnuiksi karanteenikanalassa. Broileri-isovanhempien jälkeläisten, broileriemojen kuoriduttua linnut siirretään emojen kasvatustiloille, joilla ne kasvavat 18 viikkoa.

koa. Broileriemot ovat varsinaisten broilerien vanhempia. Munittamosta broilerinmunat siirretään hautomoon. Haudonta kestää 21 vuorokautta. Kuoriduttuaan broileriuntuvikot kuljetetaan broilerinkasvatukseen erikoistuneille tiloille, joilla broilerit kasvavat noin viisi viikkoa.

Suomessa broilerituotantoa on Satakunnassa, Varsinais-Suomessa, Etelä-Pohjanmaalla ja Pirkanmaalla. Suurin osa broilerin tuotantoketjun alkuvaiheesta broilerinkasvatukseen saakka hoidetaan sopimustuotantona. Maatila tekee elintar-

FIGURE 5 A poster from the “Tipitii” exhibition.

The inefficiency of meat production raises environmental issues and finally there are issues about consumers desire for cheap food and consumer responsibility of their buying behaviour. Yet, broiler production is an important part of food industry and it offers living for numerous people directly and indirectly. Together these views - no matter of individuals own thoughts - form the whole picture of broiler production. I wanted to see how these views are present in the Internet context of the exhibition.

Internet context

A natural starting point for finding out the Internet context of the exhibition was its title. A quite surprising discovery was made about the title of the exhibition while examining online materials. The title of the exhibition reminds a song called Tipi-tii, which was Finland's representative in Eurovision Song Contest in 1962. The author of the song did question the association's right to use that title and therefore the association asked a statement from the Tekijänoikeusneuvosto (copyright council of Finland). The council's verdict was that the title of the song is not copyrighted and therefore the association could use it freely (Tekijänoikeusneuvosto, 2011). The statement is online and it is found when searching with the title of the exhibition. This gives another point of view to the exhibition. Why an exhibition about broiler production was named after the song that's message was - according the author of the song - appreciation of living nature?

When a Google search is made with a Finnish term "broilerituotanto" (broiler production in English), the first hit is the page called "Eläin koneena" ("An animal as a machine" in English). It belongs to the website of an animal rights organisation called Oikeutta eläimille (Justice for animals). The site is highly critical towards broiler production and meat industry in general. The organisation has filmed secretly pig, broiler and fur farms and published them in order to get attention to circumstances of the production animals. The second hit is a link to advertisement of Tipitii exhibition in the siipi.net, which is owned by Siipikarjaliitto, which is a parent organisation of Finnish Broiler Association. Also in the first page in the results is a discussion forum called Suomi24 with a discussion thread about broiler production. The next hit is a site called lisaalihasta.fi which is a site by Lihatiedotusyhdistys (meat information association). Lihatiedotusyhdistys is an information service by meat industry. The members of the association are Atria, HK Ruokatalo and Lihakeskusliitto ry. The site itself includes information about the environmental load of the meat production, the role of meat in nutrition and circumstances of production animals.

When the type of the search is changed to image search, the result is similar. The first image in the search result comes from the site of animal rights organisation (Figure 8). The second image is from a company that is producing and selling micro-based sewage treatment plants for example for broiler farms. The third image, a sweet baby chicken, is again from the Tipitii exhibition. In addition there are images from another animal rights organisation's site (animalia.fi), from production hall manufacturer's site (ruukki.fi) and the site dedicated to infectious

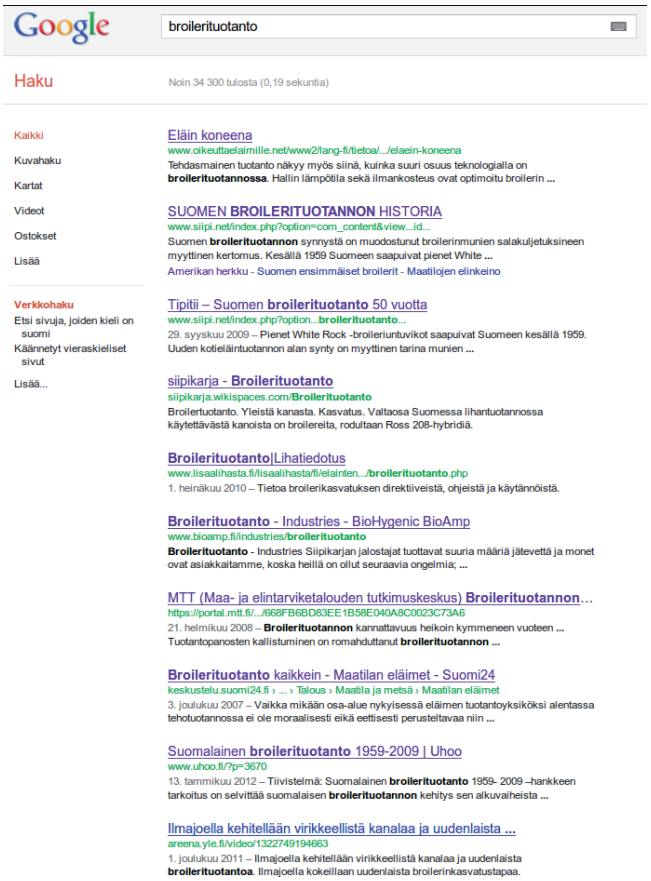


FIGURE 6 The search results with the term "broilerituotanto". (23.01.2012)

diseases that can be transmitted from animals to humans (zoonoosikeskus.fi). Animalia's pages offers links to various sources like the Report of the Scientific Committee on Animal Health and Animal Welfare, which was quoted in the beginning of this chapter. One of the images comes from a food blog called Soup-Opera. The image is from the blog article about alternative for industrial broiler production. A small farm called Viskilä is raising cockerels which grows much slower than broilers and which has more freedom. This of course is reflected to the price of the product. Finally one image presenting interior of the broiler hall is coming from a company manufacturing broiler halls.

One of the findings with the search term "broilerituotanto" was a net article called "A chicken called ROSS 508". The article is a good example of how controversial issues like broiler production can be discussed. The History of Finnish Broiler Production by Hilja Toivio has been used as one of the sources of this article. The article covers different viewing points: environmental issues, animal rights issues, consumer aspect and producer's aspect. On the other hand, the writers states that it is the consumer who decides of how broilers are produced. Demand of certain kind of broiler meat has lead to development of Ross 508, which is designed to grow fast with large breast muscles. At the end of the article there is an animation showing the killing rate of the broiler which is according to site over 140 000 broilers a day. A new broiler appears at the bottom of the page every time when a broiler is slaughtered visualising the quantity of broiler production. But it is the comment section of this article that is the most interesting. A very vivid discussion have taken place in the comments section with over 50 comments.

This article gave the key word that opened the whole spectrum of online materials related to broiler production: ROSS 508. Several discussions, blogs and sites were found when a search was made with the term "Ross 508" or "kana nimeltä ross 508". The article is linked from several discussion forums, blogs and sites including following ones:

- Kana nimeltä Ross 508 - syötkö vielä kanaa? (Anon., 2011)
- Syö mieluummin isoa eläintä - HS.fi - Ruoka (Sutela, 2011)
- Mitä ajattelin tänään: Frankensteinin perhetila (Silfverberg, 2012)
- official care instructions of ROSS 508 by the breeding company Ross (Ross Breeders Ltd, 2000)

One of the most interesting findings related to broiler production was the official care instructions of ROSS 508 by the breeding company Ross. This document offers an insider view to broiler production. Also among the findings was Anu Silfverberg's column called "Frankensteinin perhetila" in the online version of Helsingin Sanomat. In her column (Silfverberg, 2012) critisises Atria's Perhetila campaign for polishing the public image of broiler production with the term "family farm", which refers to small production units instead of large units in reality. When a search is made with the term "Atrian Perhetila", then the official site of the campaign is found but also a mention from the site of the Oikeutta Eläimille, that mentions that the organisation has made an offi-

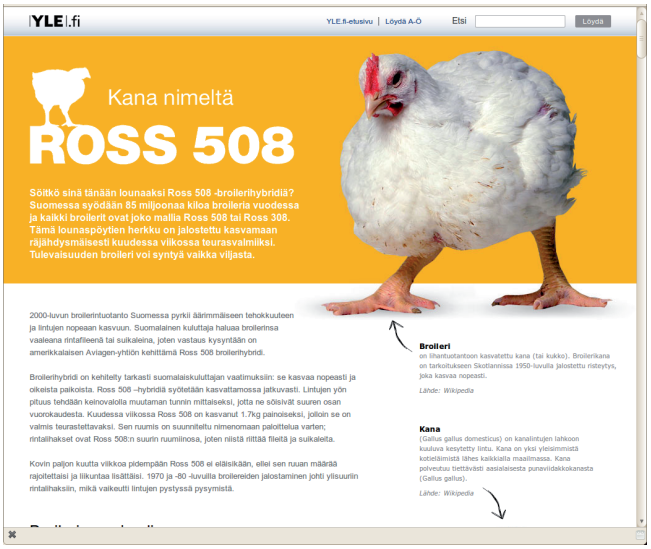


FIGURE 7 The article about broiler production by the YLE.

cial investigation request about the misleading advertising by the campaign by Atria(Oikeutta Eläimille, 2012).

The landscape of broiler production in the Internet is the landscape of information battlefield. Enormously rich net of information is revealed when searching and browsing the Internet with the search terms related to broiler production. There are emotional attacks, fact based argumentation, scientific studies, care directions, opinions and different kind of attitudes. Somewhere in the middle are ordinary people with their comments. Some of them are confused, some are upset, some are angry or defensive. As I said in the beginning of this chapter, broiler production is a controversial issue.

In the search engine universe we are exposed two opposite views of broiler production. On the other side are the meat producers and on the other side there are animal rights groups. In addition, there are vegetarians (and vegans) and people who eat meat but who do not tolerate intensive production units and excessive breeding. The animal rights organisations are dominating the Internet context of broiler production. Both Animalia and Oikeutta Eläimille are well presented in the search results with the term “broilerituotanto” (broiler production). The counter force to these sites are created by meat producers. In this context, the Tipitii exhibition is clearly in the meat production pool because its content and because of its sponsors.

What has happened is that animal rights organisations have number one rank with the search term “broilerituotanto”. When someone publishes anything concerning broiler production online, it is shown parallel to the animal rights sites in the search results and there is very little to be done to avoid this. The images of sweet baby chickens receives totally different meaning when shown parallel with images from animal rights website. In that context, the museum’s exhibition

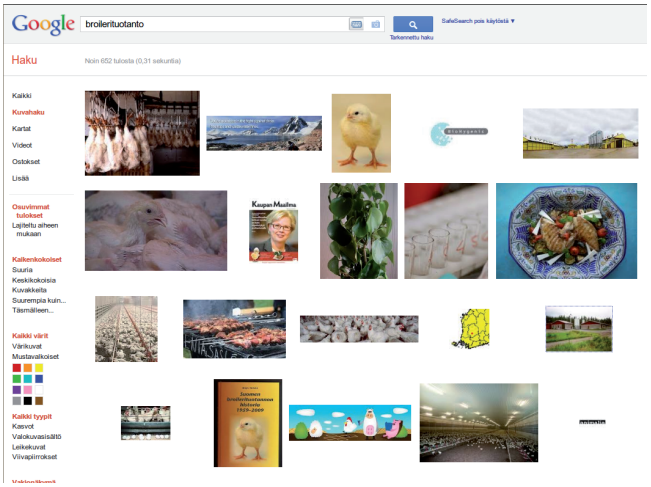


FIGURE 8 Image search with the term "broilerituotanto". The image from Tipitii exhibition (third image) is contrasted with the image from the animal rights site (first image). The query was made 6.10.2011.

appears as a PR-material that is designed to give a non-problematic impression of broiler production. If these images can be tracked back the exhibition and if they got top position in web's search engines, then there is a conflict between digital (Internet) context and the museum exhibition context.

Museum's response

I asked the Sarka Museum how they saw they role as an exhibition organiser. I send quite demanding questions about animal rights and environmental issues in general. I chose this view because it was missing from the Tipitii exhibition. Original questions and answers by the museum are in the appendix A. According the answers museum recognises its responsibility for presenting true and balanced information. However, the museum sees the the Internet as a source of uncritical and random information.

> [...] We do not have collision of interests with the maker of the exhibition or with the financier because we try to be honest and base our actions to purely objective information. We are making our exhibitions based on the right and true and comprehensive information material, not based on the uncritical and random data found on the Internet. (translated by the author)

In the light of the sponsor list and the fact that the exhibition was mainly financed by the commercial interest group related to broiler production, the museum's statement about "right and true and comprehensive material" sounds dubious. I see the situation totally differently than the museum. It is the Internet that provides independent information and alternative views and allows individuals to form their own opinions or take part to discussion. Information provided by the museum is unilateral and suspicious since it is sponsored by commercial interest group related to the content of the exhibition.

It can be argued that Tipitii exhibition was a history exhibition and therefore it was not a exhibition about broiler production in general. In other words: since the content of the exhibition is the history of broiler production, it is the history that must be at the centre of the exhibition - which is also what the museum stated when I asked if they could have issued animal rights view in their exhibition. I think this is a valid argument. But it must be also noted that broiler production has not ended but it is an ongoing industry with commercial aims and that the financing of the exhibition came from commercial interest group. Therefore the exhibition can not be treated as a documentary of the period of history that has ended but development story of industry that is currently active and that has a very strong commercial interest in influencing public opinion about broiler production.

Finally

Tipitii exhibition was not an online exhibition and the exhibition itself does not exist anymore. Still, the name of the exhibition exists online and at least one image from the exhibition can be found online. In the search engine universe the exhibition is shown together with sites criticising broiler production. When a museum decides not to tackle issues related to subject of its exhibition, will free Internet do it for them? At least in some extent yes. Different kind of views are juxtaposed by search engines without mercy. It is possible that this juxtaposition is unfair or skewed but it is there. This case raises some questions. Firstly, what is the role of the museum in this case? Should museum provide different kind of perspectives or not?

It is not fair to pin point one individual exhibition in one individual museum. However, the Tipitii case shows two aspects that can be generalised. Questions can be raised when publicly funded museum with small resources is working with commercial interest groups. The lack of resources forms a channel that exposes museums to biased - or at least unilateral - content. Not every exhibition can include every possible point of view related to the subject of the exhibition. Still, when publicly funded museum is co-operating with commercial interest groups, an extra care must be taken. If there are commercial parties involved, there is a risk that the museum lends its authority to one commercial interest group ignoring other views. There is a very thin line between an exhibition sponsored by commercial interest group and a commercial PR-material shown in the exhibition room of the museum. Museums should be very careful not to cross that line.

The second aspect is the role of the Internet. The case shows how sealed chambers of museum exhibitions are leaking. Exhibition does not start when a visitor enters to a exhibition hall and it does not end when the visitor leaves. A quite different kind of reality is waiting the visitor if the visitor decides to check what the Internet has to say about topic. The content, point of views and interpretations are escaping from exhibition organisers no matter if the exhibition is online or offline.

Is museum's role to educate, stimulate and challenge intellectually? Introduce to different ways of thinking? In this case, the stimulation and challenging is not happening in the museum but in the free Internet. It is the polyphony of the free Internet that is able show the wide spectrum of the subject. A battle field of information is exposed by the search engines. It is a fragmented view and it requires some effort to build that view but it is there. Broiler production turned out to be a complex and interesting topic that challenged one to think oneself and reflect one's relation to meat production and animals in general. All this because of free Internet.

4.5 Cases in the article "Google, world's largest museum?"

These cases are presented in the article "Google, world's largest museum?" later in this work. A brief summary is presented here. The article studies search engine visibility of digital cultural heritage and the role and possibilities of image based search. Text based searches are always more or less language dependent. The case of James Bond's toy gun illustrates this problem. The example toy gun is found from a Swedish museum. Swedish terms used by the museum serves only people who knows the terms and they meaning. Still, using images as a reference it is possible to find similar objects without language barriers.

The cases of Maretele Maimute and American Bison further illustrates the role of image based searches. The painting by William Holbrook Beard is in the collection of the Brooklyn Museum. The museum has published the image of the painting on their website. By using that image as a search criteria it was possible to find out that the painting has been used in the cover of the Romanian edition of the Will Self's book called Great Apes (Romanian title was Maretele Maimute). This finding was only possible with the image based search, since there were no mention about the cover art of the book in the Internet. Similarly, the case of American Bison exemplified a situation where link between an image and the metadata of the image could be found only with an image based search.

Image based search is able to reveal new connections between different information sources. Plain images can work as links without any explicit linking. This means that it is possible to built connections between resources by just allowing people to use images. When Internet users use images - for example from a museum's site - in their blogs or websites, these images can be found with image based search. Image based search is language independent and it does not require any actions from the users. It serves both those that use and view images and those that share images since tools for image based search are available to everyone through services like Tineye and Google's image search. But this requires that images are licensed in a way that allows them to be re-used and that the images are not destroyed with watermarks.

The case of paintings by Francisco de Goya demonstrates how unofficial resources are challenging the institutional digital heritage. Most of the images

presenting Goya's paintings on the Internet are not coming from museum sites. Many of those images comes from commercial sites, such as ownapainting.com or allposters.com, offering posters or hand painted copies of Goya's and other famous painters' works. Some of the images are from virtual collections like franciscodegoya.net or www.eeweems.com/goya/. Especially the latter site - eeweems.com/goya/ - is an excellent resource for anyone interested of Goya and his works.

Technical solutions can affect dramatically to the search engine visibility of a museum online database. The cases of The National Costumes of Finland and The Cultural Environment Portal by the National Board of Antiquities examplifies these difficulties. The content of the databases of the Craft Museum of Finland is totally invisible to search engines due to the technical implementation of the databases.

The site kansallispuvut.fi is now (2012) redesigned and the issues of search engine visibility due to the use of frames are now fixed. Google now indexes 199 pages from the kansallispuvut.fi. Also attempts to prevent downloading of the images by technical means are removed. Still, the re-use of the images is explicitly forbidden thus preventing the creation of external information sources based on the image materials from the site.

5 PATHS FOR OPEN DIGITAL HERITAGE

Problems of opening digital heritage can be divided into two groups: Problems that can be solved by technical means and problems that can not be solved by technology. We can define two views reflecting those problems based on earlier examination of different aspects and problems of digital heritage. Open Data view (technological aspects of the digital heritage) emphasis machine accessibility of the data while Free Culture view (non-technological aspects) focuses more on re-using of materials by people. The Open Data view deals the design questions of open information systems. Technical implementations have a crucial role in digital heritage and thus have that role also in any attempt to make it more open. The tools used in the field of digital heritage have limitations and in worst case the software dictates what can be done and how. Information integration - although being important - must not limit innovations of new documentation practises. Without specific knowledge of software engineering it is difficult to change tools or even demand more suitable tools.

The view for non-technical issues can be named to the Free Culture view. This approach deals with the attitudes of heritage organisation towards opening their materials, licensing, copyright issues and relation to external sources. These two views are not alternative views but they are - at least partly - inseparable. The positive attitude towards opening materials of a heritage organisation does not matter if it is not supported by organisation's information system or if they have no resources to put their attitude in practise. On the other way around: Even the most advanced open information system does not has effect if heritage organisation does not want to use it for opening their data and materials. Technology is essential part of digital heritage and it is especially important part of open digital heritage. Openness is both a technical and non-technical term in the field of digital heritage. Therefore I have tried to pay attention to these both sides in my work.

5.1 Free Culture view (Attitude)

The Free Culture view is about the role and purpose of heritage organisations and their relation to the public. Museums are seeking more visitors and more hits on their web pages. Are they willing to see themselves as an information providers for external information networks? Networks that might not (at least not directly) bring more visitors or page hits? Are they able to accept that they are not the only authors of cultural heritage? Do they support permission culture or free culture?

There are three non-technical issues that concerns open digital heritage in heritage organisation level. First is the attitude. Are the content holders willing to release their materials (or at least part of it) and giving up control? The second issue is how materials are licensed. Merely putting materials online does not yet create an open digital heritage resource. The third issue is the resource question: Does the content holder know how to publish their materials online in a way that makes materials accessible?

> While this requires museums to deliberately lessen their control as to how their objects will be used, discussed and contectualized, the rewards are only beginning to be imagined. Deeper levels of trust and collaboration with users could not only improve learning and increase audience engagement, but also enhance knowledge and stimulate creativity across the board. (MacArthur, 2007)

In the framework of this study I have tried to bring together different aspects that affect the openness and usefulness of the digital heritage. For example, there are no technical solutions for copyright problems but there are different approaches in order to cope with copyright issues. Or, there is no technical solution for the situation where museum's funding is partly based on visitor count of their web pages and therefore museum is reluctant to open their material to be used outside. Free Culture view is about attitude. Do heritage organisations see themselves as bazaars where everyone can pick materials according one's interests or do they see themselves as temples with harmonic and controlled information. These are the questions that can be solved only by heritage institutions themselves. In this work I could only point out these issues for those who are interested to see them.

Signs of Free Culture view are emerging. Materials are 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. Museums publishing their images in Flickr commons, Brooklyn Museum's freely licensed images, and The Yale Digital Commons are just some examples. The most advanced project in the direction of Free Culture view is Smithsonian Commons. According to the new media strategy of Smithsonian Institution, the purpose of Smithsonian Commons is:

> Publish content so that it can be shared, tagged, collected, rated, re-purposed, commented on, and exported to other Web sites. Optimize Smithsonian digital content with meta data and semantic structure to maximize availability to search engines. (Smithsonian Institute, 2009)

Smithsonian Commons is not reality yet. But I am convinced that when established, it will have an effect on museum community. Smithsonian Commons is an open invitation to participate, while, for example, Europeana is an information portal. The difference is in the attitude and the presentation of that attitude. Smithsonian Commons is targeted directly to individual users with explicit and free licensing. Europeana, in its current form, is more like a virtual display case that allows watching but not touching. Because Europeana does not aggregate the content but is only linking to the content, it is difficult to find out what users are able to do with materials found. However, this problem is addressed in the Europeana Licensing Framework:

> Users need good and reliable information about what they may do with it. Whether they can freely re-use it for their educational, creative or even commercial projects or not. The Europeana Licensing Framework therefore asks data providers to provide structured rights information in the metadata they provide about the content that is accessible through Europeana. (Europeana, 2011b)

Until Europeana is able to provide proper licensing information and search facilities based on licence, it not serving individual users best possible way in their path in building new heritage.

5.2 Open data view (Tools)

In this chapter I will discuss about design principles for open information systems in museums needed for open data approach. Although I am arguing that the most important factor in "openness" is the attitude, the openness in digital networks is still executed by technical means. In other words, openness is not a technical issue but it has consequences that affects to the technical implementations. There are three major factors that are critical for information systems used in open digital heritage field: openness, flexible (and preferably semantic) documentation model, and built-in long-term preservation. Openness is - quite obviously - the key value in open digital heritage systems. This does not necessarily mean openness of the source code of the applications used although open source software fits well to the concept of open digital heritage. More important than the openness of the code, is the open machine access which in turn is made possible by using open formats and open interfaces. Different kind of large-scale re-use - like mashups - is only possible with machine access.

In the open data model, the museum provides interface where users can fetch data. In its simplest form this means a HTML-files where information can be scraped and in the other end there is a full API and access to semantically explicit XML files. In this model museum has still clients but they are not very loyal to the museum. The museum is just one information provider for example for mashup. Open data approach means that by providing an open, machine accessible information system, a heritage organisation accepts their role in the

open network of information. Machine access might not be directly meaningful for the most of the people, but it opens doors for system builders.

The documentation model is an important factor of creating digital cultural heritage information. If software used in documentation forces to think through tangible objects, then the software is already restricting the documentation process. Flexible documentation model means that software should respect the views of the documentation makers. But in deep level it should keep documentation commensurable. This is only possible with using a semantic glue between different kind of documentation approaches.

The third necessary aspect of open information systems is long-term preservation. I am proposing a built-in long-term preservation approach which is based on open access of semantically-aware XML files. The idea is that a repository of XML files can be mirrored by a third parties without any resources needed from the provider of those files. This model is especially suitable for small organisation or individuals. This is quite opposite model compared, for example, to the Finnish KDK-project, where organisations are responsible for delivering long-term preservation packets (Kansallinen Digitaalinen Kirjasto, 2011). In following chapters I explain the characteristics of open cultural information systems.

Distributed architecture

The central question of the openness of information system architecture in the field cultural heritage is how information is stored. Current systems are mostly database driven. Many of the problems of openness and flexibility of the systems are related to this database centric information storage model. Another aspect related to the information systems in digital heritage is how the rest of the system is built around the information storage.

There are two design principles when it comes to organising software architecture. There can be a one, monolithic software that has all the functionality needed in it or there can be several independent programs that can be used together in order to accomplish the task in question. Microsoft Word or OpenOffice Writer are examples of the first case. With one application it is possible to write text, format it, check typing errors and print it. An example of latter case is the Unix philosophy: "Make each program do one thing well. To do a new job, build afresh rather than complicate old programs by adding new feature." Between these models is a plugin-based solutions. In the plugin model there is a central software that provides an application programming interface (API) for plugins. With plugins it is possible to extend the functionality of the software. An example of plugin-based approach is the Firefox browser and its add-ons.

Monolithic approach has some benefits and drawbacks when compared to distributed model in a heritage use. One program is easier to manage than several individual programs which saves resources of the organisation. Also, a monolithic software is provided by one supplier, which simplifies technical support. Monolithic programs usually provides unified user experience through unified user interface, although it must be noted that this is a benefit only if the user in-

terface is well designed. Nevertheless, the monolithic architecture has also some serious drawbacks.

The monolithic architecture is rigid which makes modifications difficult even if there were a source code available. The code base is large and modifications are difficult. The workflow, the user interface and the data structure are modelled according to some assumptions and then built in to the software. This might force users to adapt the structures of the software instead of supporting the structures and workflow of the users. In the museum field this could be problematic since individual museums have their own established workflows and data modelling schemas. On the other hand, harmonising the work flow might have good consequences since this affects positively to the formation of the data. Another problem with monolithic architecture design is its relation to external world. The system is built to "take care of everything" which means that the system is by its design self-centered. There is no need to communicate with the external world and alternative workflows are not supported.

Although monolithic systems are self-centered systems by nature, they can provide API and export functions, which makes them more open. With API it is possible to extend the functionality of the monolithic architecture. The nature of the API defines what kind of tasks can be accomplished with it. The API could be a simple query API which allows querying of a database, for example. Or API can be a full interface that allows creation of applications that could replace some of the function of the original software.

The distributed model with several independent applications might be more complex to maintain than monolithic software. Having different application means that there might be several suppliers involved which makes technical support more difficult. However, there are also some major benefits in the distributed model. First benefit of the distributed model is that the distributed model must have a way to communicate with other programs and this communication method must be properly documented. If there are several applications that share, for example, a database, then it is possible to write more applications sharing the database. In this sense, openness is built-in in the distributed model.

A client-server model is one way to implement this communication between different applications. A central application is called a server and others work as clients. For example a database application can be a server and other applications connects to the server when they need to query the database. The server provides a documented API that can be used for querying. At least in theory, this kind of system is quite flexible since new applications can be added without modifying the others. However, as I stated earlier, APIs also have some issues. What can be done with the API is defined by the designer of the API. Secondly, API can be used as a control system and if the API changes, that will break all the applications that are using the API. Most of these problems can be avoided if we switch from databases to file-based architecture.

File-based system architecture and open file formats

The idea that I am proposing for open information system is that data is stored in semantically marked, atomic text files (preferably XML-based) without a need of relational database. Basically files are just ordinary text files that can be opened in any platform. In its simplest form file-based system can be constructed from hand edited HTML-files. Although this approach lacks of formal semantics and updating requires direct editing of files, it is still much more open than if the same information was stored for example in a local Microsoft Access-based database. It is also much more flexible since HTML does not set restrictions to the organisation of the content. Still, if files have uniform base structure, the content can be harvested by external actors. More advanced solution can be based on XML files and formal semantics like CIDOC-CRM (see IDA-XML).

For information retrieval purposes a database or different kind of aggregation files can be created. The purpose of the database is to offer a query interface for the information storage. This means that database can be destroyed at any time since it can be re-created from the XML files. Aggregation files are automatically constructed from source files and they can offer list of all files available or for example or all person names in one file. This makes it easier for developers to use materials.

Important part of this design is the overall openness of the system. What can be done with the materials depends only skills and imagination of developers. However, the open nature of the system also introduces a new problem. Although most of the information in the cultural heritage system can be regarded as a free information, there is also information that cannot be published or information that is not wise to publish. Information that falls in the first category is everything that is protected by laws like privacy laws which limits the use of collected personal information. In the second category - the information that is not wise to publish - there are for example an estimated monetary value of the objects or insurance information. In distributed model this problem can be solved by having separate data repositories for open data and non-open data (see figure 9). It is good to note that having separate data repositories for open and non-open does not mean having separate applications for managing data.

Files of the file-based system form a long-term solution for permanent data storage. Files can copied without any problems at any time. The essential part of this design is that system itself is always ready for replication and long-term preservation without any explicit export activities. Data files can be mirrored easily with simple file syncing methods and at the same time, the mirrored data is suitable for long-term preservation.

Universal and permanent identifiers

Having open data type access via file-based system itself does not guarantee that these files are permanently available. When museums have any materials online, it is important to realise that they are then part of open information network.

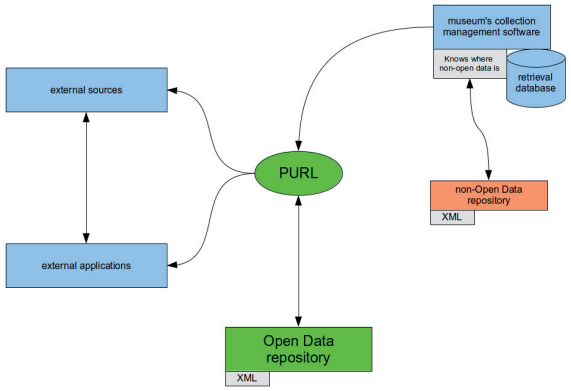


FIGURE 9 The role of persistent URLs in open information system.

Links to the museum's site or database bring visitors but even more important is the information that those links provide to outsiders. Therefore it essential that museums - or heritage organisations - do not break that information network. Heritage organisations should offer only permanent addresses to their materials.

Persistent URL means an identifier that is unique, permanent and accessible. For example relational databases uses identifiers internally. In certain sense those are permanent but they are not unique (when compared to other databases) nor are they accessible (since they are internal). In the museum context a natural base for URIs are the collection of item identifiers. These identifiers per se are not universal resource identifiers but if they are combined with unique domain name, they can be used as such. The Internet domain of the museum can be combined with the identifier of the collection object like this: www.somemuseum.com/234654. This is now a unique identifier that can be referred from outside.

The idea of permanent URL is that it provides a universal and implementation agnostic way to refer to something. Currently when we want refer to an online database, we often must use a technical URL that is provided by the application. A technical URL includes references to application's internal structure and therefore URLs are not persistent. If the application is replaced with other type of application, this change will break all the links - internal and external - to the database items.

Permanent URLs can be used internally in the information systems in order to combine different parts of information related to that specific URL. This makes it possible to use distributed architecture for the information system.

5.3 Hacking Digital Heritage

Participating to the creation of digital heritage covers all actions from the creation of high-tech computer-based exhibition in the exhibition room of the museum to the using of freely available image of an artwork in one's blog as a background image. The scale of different actions is wide and it is not possible to cover all these actions here. In the following chapters I will discuss about one activity: the possibility of hacking of digital heritage.

The combination of museums and hackers is interesting and challenging. In many ways, hacker and museum cultures seem to be quite opposite cultures. The hacker culture is playful and non-hierarchical, the museum culture is more serious and often has some organisational stiffness. Hackers start they actions without permissions or bureaucracy, in the museum field projects starts with careful planning and with numerous meetings.

> For the hackers, the typical governmental way of having endless meetings, forming countless committees, drafting tedious strategy papers, and so on before anything happens is at least as great a pain as doing market research to justify an idea before you can start to create. (Himanen, 2001)

Even if that there seems to be a contradiction between organisational bureaucracy of museums and the playfulness of hacker culture, it is also possible to see some possibilities. From the point of view of the hackers, the museums have interesting digital content, good devices and they can offer public space for different kind of experiments. The hackers can provide to the museums for example crowdsourcing platforms, data mining solutions, computer-based exhibits, expertise on hardware and totally unconventional ideas.

Creation of new services, applications and experiments from existing heritage materials requires a public digital access to these resources. A public interface provided by a heritage institution allows experimenting with the materials and making application prototypes with real data from anywhere in the networked world.

When an organisation opens a public, digital interface, it is an invitation to create new uses of material. It also means that the service provider's control of the material is loosened. Material flows "somewhere out there" without possibility to control the context where material is used. It is important to note that opening digital interface and inviting third parties to create new ways to use materials differs radically from crowdsourcing and user-generated content approaches that many museums have experience. Crowdsourcing involves a very specific task and it is good for solving simple, but numerous tasks that are difficult to computers. It can be used, for example, for correcting mistakes made by optical character recognition software in digital scans of old texts. User-generated content, on the other hand, involves collecting content form users in form of comments or via social media. The difference here lies in the role of the museum. In crowdsourcing and user-generated content approaches museum is the main participant that defines the framework of activity. Opening digital interface works differently.

Museum's role in a pure Open Data approach is the role of service provider and the role of being an information source.

Outside Institutions

The Open Data movement uses the term "data journalism" for the process of making open data sets to accessible to the wide audience. This includes finding an approach to the data, harvesting the data and programming, and possibly visualising the result. This schema is possible also with open digital heritage. However, like I stated earlier, the difference between the open data and open digital heritage is that modifying of the content from open data source is meaningless while modifying the content of heritage source is one form of contribution to the digital heritage.

The Open Data style access makes it possible to built metamuseums or metacollections that are based on institutional collection materials. For example, I am interested of bicycles and history of them. Now, I want to built a site dedicated to the history of cycling in Finland. First, I would fetch a list of open data information sources that are provided by the museums. Then I study programmatically what concepts, key words and texts are found on those sources and see if there are any mentions of bicycles or cycling in general. Then I could made a compilation of founded texts and images, reorganise them and add more information. By adding a system where others could insert their own images, experiences and memories, I would have created a metamuseum of bicycles that is more than just a combination of different information sources. This metamuseum is an extension of an existing collections but it is created totally independent from the institutional material providers. It can point back to the original source so that users can see also the original data.

In the previous example, a user created a totally independent metamuseum that was based solely to the needs of the user. Nevertheless, we might think of a different kind of scenario where the provider of the freely usable materials is hoping to receive direct results for their own use. For example, a museum might want to find all the images from their collection where there are people in them, which in turn would help their staff to identify these persons. At least a partial solution could be provided by using a computer vision techniques to detect faces from the images. Since there are thousands of these images, there must a be some kind of automation involved. The musem has no funds or expertise for this kind of project.

In a another scenario, the heritage institution wants to combine the Open Data approach and crowdsourcing. In other words, they publish open data but they also want to set a task to be fulfilled by crowsourcing and they wish that the platform for the crowdsourcing could be created by volunteers. As an example, we can imagine a situation where a museum has interviewed several people about some collection objects and asked them share their memories and knowledge related to these items. These sessions are record with a digital video camera. This kind of documentation is fast and quite effortless. At its best, this kind of

documentation is able to produce very natural and contextualised information. The problem is to how to find information about certain item from a long video without a need to watch the entire video. The second problem is how to connect information in the videos to the existing collection data about these same items. One solution is to publish videos online and ask help from the Internet community. Since participants may not want their appearance to be shown to the whole Internet, it might be easier to create videos so that only the objects and the hands of the people are visible but never the faces of the participants. Hands are important since people usually like point out the details they are talking about. This increases the information value of the video.

The task in this case is to make a tool for video-based documentation that can be further used in a crowdsourcing project in order to actually tag video materials. One possible workflow would be as following. Individual objects are tagged from the video so that they can be easily found from the video and further, the tags can be linked to the existing object identifiers. If there is no earlier digital documentation and therefore such identifiers, then the tags could provide a starting point for that documentation. Spoken information can be then extracted from the video to a more formal form by volunteers.

I designed a simple video tagging application in order to demonstrate this crowdsourcing attempt and in order to test the information system design principles for open digital heritage. The application called IDA.video is presented in more detail in the Technical Proposals section.

A quite natural way to organise this kind of Open Data hacker invitation is to set up a competition. Like I stated earlier, Open Data access does not allow setting any task or limitations of use by the provider organisation. The data is provided for free use without any conditions. However, competitions have rules and this gives a framework where quite specific requirements can be set without interfering the principles of Open Data.

There are certain aspects that the organizers should take account in this case. First, they must realise that they are dealing with a very heterogeneous group of people with varying skills, attitudes and life situations. Someone wants to work alone and some other prefers working in groups. The other wants to use technology X while others want to use technology Y. Therefore the initial assignment should be flexible when it comes to organisation of the projects and technology used. The institution should also make sure that the task can be accomplished by working remotely. This increases the number of potential participants. In a practical level, the organizers need to make sure that they receive the source codes of the applications with a free licence. Otherwise they might end up to vendor lock situation, where no one is able to alter, repair or update the applications.

Organizers should also understand some aspects of software development. It is possible that a novice programmer or programmers are not able deliver acceptable solution despite all the effort. The end result is then an application that does not work properly or that is highly unreliable. If the original idea was to use this software for crowdsourcing, it means that the users become frustrated and the crowsourcing effort is not fulfilled. Repairing the software that may not be

well documented, or that is designed poorly from the beginning, requires considerably amount of work. In addition, for a hacker this kind of "fixing" is rarely as pleasing than creating something completely new.

Another problem scenario is that the development of the software stops for some reason when there is, for example, only ninety percent of the required functionality implemented. This "almost finished" situation is difficult. Although there is only ten percent of functionality missing, it does not mean that there is only ten percent of coding work left. This ten percent might require very specific expertise, like skills in statistical analysis, for example. Finding a person that has required expertise and who is able to complete the application can be very challenging.

There a couple of issues that are specific to open data crowdsourcing. Like I mentioned earlier, a broken crowdsourcing platform efficiently destroys a crowsourcing project. The second problem is a user interface design. Easy and attaractive user interface is essential for a crowdsourcing project. However, designing this kind of interface requires other skills than coding. The third issue is defining the task itself. The task size should be reasonable. In traditional crowdsourcing projects the task can be identifying of millions of hand written words, for example. The size of the task does not matter since the work can be shared between participants and the actions of participants are independent of each other. The situation is different in a software project that aims to a creation of a crowdsourcing platform. There must be a consensus of the methods and division of work between the participants. When the task is complex, the possibility that the consensus is not found and the division of work is not possible.

Opening information and documents for free use benefits individual enthusiasts as well as random heritage seeker. However, if heritage institutions want to benefit from hacker community, they must prepare to use some resources to it. They must define what they want, they must have understanding on open source software development, they must offer a channel for discussion and questions and they must have some understanding of the concepts involved. Still, there is no guarantee that there will be any usefull input. However, there is a possibility for a very successful project.

Inside Institutions

It is a sad fact that small museums have no resources to create fascinating, interactive applications for their exhibitions. One can even speak about digital divide between small museums and large, more resourceful museums. However, at least part of these resources could be available also for small museums if museums were hackable. The idea of "hackable museum" means that a museum opens its materials for third parties through a public application interface enabling programming communities to create new uses and new ways to present materials. Could museums benefit of the hacker community for creating computer-based exhibits and what would that require from the museum?

Computers and skills needed to use them efficiently have always been a

problem for museums (Parry, 2009). According the survey carried out by Open Exhibits4 in 2010 (OpenExhibit, 2010), 54.7 percent of all respondents said that the main reason for not having computer-based exhibits was the lack of technical staff or staff time. However, 77.3 percent was interested or very interested to develop their abilities to produce computer-interactivities in house. In the same query executed in 2009, 88.7 percent of all respondents were interested or very interested to use open-source software for development of computer-based exhibits.

Museums except that computer-based exhibits are something else than mouse controlled info kiosks. There must be an intuitive user-interface, it must be interesting and the “wov-effect” wouldn’t hurt either. One of the respondents said:

> "it needs to be different than what a user could find at home or the library. needs to be transparent in its user interface. needs to represent an idea or phenomonon in a way that uses the medium intelligently".

Creating new kind of user-interfaces requires highly specific skills like computer vision programming and knowledge of micro-controllers. In free markets the monetary value of these skills can be quite high and therefore purchasing them as services is not an option for small museums. Other option is to purchase on-the-self products - which may not be affordable neither - but even they requires technical skills and time in order to set up properly. Still, museums are very interested of creating computer-based exhibits and museums have a positive attitude towards open-source software. When these two are summed up, it seems that museums and hackers could have a future together.

Museums could be interesting targets for programming communities for three reasons. Firstly, museums can offer interesting problems to hackers: Developing new kind of interfaces based on computer vision or using geolocation information in order to guide the visitors, for example. Secondly, museums do have interesting digital content. They have huge amount of digital image data, 3D-models, digital audio, video materials, text files and databases. This is suitable material for image analysis, virtual reality applications, data mining or something to be combined with information from other open interfaces like Wikipedia. The third thing that museums can offer is a public place – and therefore publicity – for experiments.

However, even if hackers could see museums as places for interesting projects there are at least two major problems in a hackable museum. The first one is how are the needs of the museum adapted to the installation designed by an independent designer or designers? Obviously museums are titled to define what kind of exhibits they have in their museum and what purpose they serve. The problem is how willingly are volunteers to respond to these needs. If museum sets too high requirements for example in pedagogical sense for their hacker-generated exhibits, there might be no exhibit at all. What interest hackers is the technical challenge and possibility to create something fun that others can enjoy. If there are too much other requirements that must be also taken account, it is possible that these requirements takes the fun out of the task. This obviously is a prob-

lem from museum's point of view. Museum is not interested about the technical solution but what can be achieved with it.

At least a partial solution can be found from open-ended play which was also mentioned in the survey by Open Exhibits. By open-ended play I mean that the system has no path to follow and therefore the outcome could be different for individual users. Open-ended play gives room for exploring, browsing and wild combinations. It is also much easier to achieve when there is little interaction between museums stuff and designers of computer-interactivity (hackers in this case). It also easier for museums to define what they want if the end-result is not fixed.

The second problem is the maintenance of the computer-based exhibits that can contain highly tailored elements. Event though the hardware setup is simple, there might be complex dependencies between applications, some sort of calibrations that must be run, or hardware dependent code that breaks the system if one component have to be changed due to hardware failure. The situation is similar to complex installations by media artists. There is no simple solution for the maintenance problem. Museums may try persuade developers to write some documentation, which is not a very popular task among developers. On the other hand, developers of open-source software has positive attitude towards error reports (bugs). Bug reporting is one way to improve the software and this might be also the maintenance route in this case. It means that museums have to know something about the practises of bug reporting.

In the Technical Proposals chapter I introduce an exhibition application called TUI.maps. The purpose of the application was to offer information about the buildings of a certain area in a easily accessible form. It was a combined experiment of using a tangible user interface and testing the API provided by the IDA platform. The point here is that the application could have been created anywhere in the world and it could be also used anywhere in the world if required hardware was available.

Open digital access and hacking competitions could be an interesting model for museums. After opening a public interface, museum can organise hacking competitions or "exhibition hack labs" for programming communities for creating new ways of presenting their materials. This would help them to make their materials more accessible and have interesting exhibition applications for their visitors. However, the lack of generic interface makes it difficult to participate since every museum has their own methods of accessing data. In order to invite hackers to the museum field, it is important that there is digital access to the museum's materials without bureaucracy and interference by the museum.

6 TECHNICAL PROPOSALS

My main methodology for finding design principles for open information system for heritage use has been prototyping (Häyrinen, 2012). I wanted to have a very practical, firsthand experience of technical aspects of digital heritage and use that experience in order to find some design principles for open digital heritage information systems. I have developed software prototypes in order to find practical solutions for open and sustainable information systems. These experiments are not finished applications. These experiments include a cultural heritage database, a presentation software, a video tagging demo and tangible user interface experiment for an exhibition use. The first version of the database was implemented with a traditional database system and the second experiment was done using XML as a storage format.

6.1 CIDOC-CRM and semantic templates

While searching solutions for problems in data modelling during the 3D-Bridge project, I found the Conceptual Reference Model (CRM) developed by CIDOC. CRM provides an even-centric model for documentation thus replacing item-centric approach. In addition, the CRM is developed for cultural historical documentation and it is easily extensible. There was no enough time to adapt CRM in the 3D-Bridge project but I chose CRM as a basis for my own documentation framework. The framework was called IDA-framework. It was developed in co-operation with the University Museum of Jyväskylä. The framework implemented CRM as a documentation model and it introduced the concept of semantic template for managing the input for different kind of targets. I have presented the IDA-framework more detail in my conference papers in CIDOC 2008 and EJC 2010. These papers can be found on the Included Papers section.

The initial idea of the IDA-framework was to create a test environment for non-item centric documentation of cultural heritage. CIDOC-CRM offered an event-based model that could replace the item-centric model and that also could

provide the semantic basis for whole documentation. IDA-framework used the concept of semantic templates. These templates were modelled according the principles of CIDOC-CRM:

> The idea of a semantic documentation template is to provide a record type specific documentation frame that can be defined by organisation responsible of documentation. The template defines a typical case for a record including default properties, required events, possible events and possible parts. More precisely, documentation template maps a part of thesauri to a partial domain ontology that is build on top of CIDOC-CRM. (Häyrinen, 2011)

A semantic template is a template for a documentation of certain specific group of targets. This target can be a person, a building, an event or certain type of furniture, for example.

6.2 IDA, a Relational Database-based Prototype

The CIDOC-CRM consists of classes and properties. The class hierarchy and properties are held in RDFS files. These files also includes also additional classes that expand the classes of CIDOC-CRM. Actual records (eg. instances of classes) are linked to this hierarchy with many-to-many relations stored in separate join tables. Properties that links indivual records to other records - like, for example, a property between a person and the name of the person - are stored in their own table.

The actual data, like dates, appellations and descriptions, are stored in data tables, which are then linked to the records via join tables. This means that a record in IDA is merely an identifier number. This, very atomic structure allows very detailed referencing to different parts of the documentation. For example, the title of an artwork is an entity that can referenced independently from the artwork itself or the construction event of a building can be documented as a separate unit.

The IDA-framework has a simple XML-based application programming interface (API) that can be used for adding, modifying and querying the content. The query functions are open for everyone but adding and modifying functions requires an approriate account.

The listing 6.1 shows an example query. The first node is the command node that defines the functionality of the query, which is searching in this case. The second node is the class of digital images which is followed a property “represents” and the “Buildings” node is the object of that property. If the “Buildings” node had on identifier (id), then the query would search images that are representing certain building. Finally, the “result” node defines what information is send back. In this case the result includes the filename of the file and the name of the building or buildings that image presents. The idea of this very verbal query language is that queries are be self-documented and understandable also for outsiders. If a

LISTING 6.1 An IDA query for images presenting buildings.

```
<?xml version="1.0" encoding="UTF-8" ?>
<search>
  <Digital_Image>
    <represents>
      <Building />
    </represents>
 </Digital_Image>
 <result is_identified_by="filename">
    <represents is_identified_by="name" />
 </result>
</search>
```

query language is simple, then using it is possible for more wider audience than more complicated query languages.

The development of IDA revealed some limitations of relational databases. In order to build flexible documentation system, a relatively complicated structures had to be used. The SQL queries themselves get longer and longer to the point where it was hard to figure out what each query actually did. As a result, the complexity of the system increased accordingly to the flexibility of the documentation model. This was because of the relational database model itself. Relational database model is predefined and rigid and every attempt to make an application more flexible increases the complexity of the system because every exception must be coded against the nature of relational database.

The most serious problem of relational databases is their suitability for long-term preservation. Relational databases are built for effiency in mind for handling large datasets. This requires different kind of optimatisations in a way the data is stored. This means that data is stored in application specific form that can not be accessed without the specific database management application which naturally forms a problem for long-term preservation.

Relational database introduced also other kind of problems. Although SQL is a standardised, there are different SQL dialects between DBRMS which are not compatible with each other. In other words, switching to a different DBRMS - like switching from MYSQL to POSTGRESQL - is not trivial and requires code changes. This means that some kind of vendor lock is created. However, there are intermediate layers that solves this problem by translating queries to the DBRMS-specific dialect. One of these layers is MDB2, which was used also with IDA-framework. But even using this kind of layer did not solve all the problems. There was even difficulties when application was moved from development machine to the production machine. While everything was working in the development machine, the system did not work on production server because there was a different version of MDB2. If even I - the creator of the software - had problems to get the system to run in different environments, how could that work with less experienced users?

The development of IDA revealed several critical points of software development in the field of cultural heritage. The first notion was that discussion with museum staff who has no IT-expertise, gets centered around user interface issues.

A user interface is the only visible component of the software and therefore this is not surprising. The user interface is usually the only contact point for museum workers to the software and for them, it is a tool that they use on their daily work. One can say that for an average museum worker the user interface is equal to the application. This creates a difficult situation for the software development. For example, there can be two totally different kind approaches or systems (for example a database driven system and XML-based system) that lead to very different kind of solutions in implementation and software architecture levels. But for the users the only thing that matters is the user interface. This makes communication difficult between developers and users.

The IDA-framework was tested with the materials of University Museum of Jyväskylä. The material included maps, blueprints of the buildings and images of the Seminaarinäki alue. The most challenging part of the process was the ontological analysis of the materials and defining semantic templates according to that analysis. The software was developed simultaneously, which made things even harder.

6.3 IDAXML, an XML-based prototype

The development of database-driven IDA showed some shortcomings of relational databases in the field of cultural historical documentation. The main issues were the data modelling issues related to the rigid data structure, the problems in opening the data and suitability for long-term preservation. Therefore an alternative approach was needed. This approach is based on using XML files as a main data storage.

IDAXML is a partial re-implementation of IDA-framework with XML backend. The idea was to test file-based approach as a replacement of relational database. IDAXML is based on CIDOC-CRM and it uses semantic templates similar to IDA. IDAXML is a XML file repository with an API that allows addition of the content (currently modifying is not implemented). The main difference from the previous, database-driven approach is that data is stored in XML files that can be accessed directly by anyone. IDAXML has no internal data structure like relational databases. Instead, IDAXML exposes all data to external inspection.

When a IDAXML receives a new record via its API, it is first validated against the structures of CIDOC-CRM. Then each the record and each of its properties are given unique identifiers. Finally the actual XML files are written. Each record has its own XML file which includes all its properties. For example, any man-made object has at least a production event. Therefore each record of man-made object has at least to files: the record itself and the production event of it. They both have individual ids that are expressed also in the file names. The file name of a XML file includes also a class identifier of the record.

The listing 6.2 shows a person record. Appellations are included in a same file than the record that they label, but they still receive they own identifiers.

LISTING 6.2 A person record with an appellation property and birth event.

```
<?xml version="1.0" encoding="UTF-8"?>
<Person id="E21_1jgAS">
  <P1F.is_identified_by>
    <E41.Appellation id="E41_8hCck">
      <has_value>
        <name>Ton Roosendaal</name>
      </has_value>
    </E41.Appellation>
  </P1F.is_identified_by>
  <P98B.was_born id="E67_zXxfJ"/>
</Person>
```

However, events are stored always as individual records. Therefore the birth event of the person has its own identifier (E67_zXxfJ) and its own XML file called "E67_zXxfJ.xml". All properties are bidirectional which means that the birth event file includes reversed property to the person record.

Using atomic XML files has certain benefits. Firstly, using several individual files instead of one gigantic file makes up-to-date cloning (mirroring) of the content simple. Instead of downloading one big file every time the content changes, one can fetch only files that have changed and files that have been added. Second benefit is that failure in file IO when saving the content does not corrupt whole information storage but only one file. Thirdly, separate files per identifier makes id-based access very fast and trivial. Every record can be referenced by the file name. This means that harvesting the data can be done without the need of an API.

IDAXML does not provide any query functions for the content. Instead individual XML files are offered for direct online access. This kind of structure is not very efficient for information retrieval purposes. Although records can be fetched instantly by their id, it is not easy to search, for example, persons by their names. There are two possible solutions for this: use of aggregation files or use of relational database for information retrieval. The idea of aggregation file is to collect certain kind of information to a same file. Typical case would be names of persons, names of places, titles of artefacts etc. This way searches can executed simply by examining specific aggregation file instead of thousands individual files.

The creation of aggregation files is technically simple. The most simple case would be the creation of an aggregation file that includes every record in the system. This allows using more efficient XML tools - like XPath, XQuery or XSLT - for querying the content. The listing 3 presents an XSLT file that displays all persons and their birth years sorted by person names from the aggregation file from IDAXML. In a real production environment this aggregation could be done in a separate background process.

The second option for solving the query problem is the creation of database for information retrieval purposes. This makes querying very fast and familiar for those that are used to work with relational databases and SQL. However, this approach has some drawbacks. Firstly, creating an efficient database is more

LISTING 6.3 An XSLT file for displaying person and their birth years.

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

<xsl:output method="html"/>
<xsl:key name="nameById" match="/root/E41.Appellation/P1B.identifies" use="@id"/>
<xsl:key name="birthById" match="/root/Birth/P98F.brought_into_life" use="@id"/>
<xsl:key name="dateById" match="/root/E52.Time-Span/P4B.is_time-span_of" use="@id"/>

<xsl:template match="/">
<html>
<head>
<link rel="stylesheet" type="text/css" href="my_style.css"/>
</head>
        <h1>Persons</h1>
        <xsl:for-each select="root/Person">
                <xsl:sort select=".//name"/>
                <xsl:variable name="personId" select="@id"/>            <h2><xsl:value-
                    of select=".//name"/></h2>
                <xsl:for-each select="key('birthById', $personId)/..">
                        <p><xsl:value-of select=".//start_year"/></p>
                </xsl:for-each>
        </xsl:for-each>
</html> </xsl:template>
</xsl:stylesheet>
```

complex than creation of aggregation files. Secondly, the database requires more maintaining since it must also reflect possible changes in the structure of the content. Thirdly, a plain database is not usually enough but there must some kind of high-level API instead of plain SQL interface.

These methods for improving information retrieval can be provided by the content provider. However, an important notion is that both of these methods can be performed by outsiders. They can create aggregation files privately or publicly and they can create databases by harvesting the XML files. External developers can even provide independent API for the content for everyone to use. This openness has two effects. The first effect is that this saves resources of the content provider organisation. Although it can provide more efficient access with the methods described earlier, it does not have to do it. The content is still available even if the organisation does not make any effort in this front. The second effect is that developers have full freedom to organise and then query data as they wish.

Having a XML file-based storage means isolation between the internal tools of a heritage organisation and accessing methods for the content they provide. The organisation is free to modify and change their internal tools. These changes are not reflected to the content storage. This makes the system open by its design. Using XML files directly is not simple but it allows outsiders to build functionalities with the data, like a query API for example. Although IDAXML is more a demo of a open data design principle for digital heritage than a finished application, it was still able to demonstrate the benefits of XML-based approach. Using XML files as a data storage solved many issues with databases. XML files can be accessed as they are which makes them open data. The whole structure is flexible which matches well with the needs of the cultural heritage documentation.

Despite the flexibility, the system provides a permanent environment for managing cultural historical data. Text-based files can be harvested in many ways and structures can be revealed even without any further documentation.

The main problem is still semantic modelling which remains as a difficult task. Other downside of this approach is huge amount of individual files. Also, since search engines do not index XML files, an extra effort is needed to ensure search engine visibility. However, a search engine friendly HTML presentation could be quite easily constructed automatically from XML files.

One interesting possibility that could solve some of the problems in this kind of systems is text encoding. In text encoding interesting text elements in a free text are marked with tags. One example of this kind of markup is Text Encoding Initiative (TEI). For example, persons and names of the places can be marked in the text. This allows reliable identification of persons also in cases when persons are identified indirectly like using word "he" or "Mrs. M". This allows documentation with free texts still enabling formal markups. For cultural historical targets this approach suites well. A good example of successful text encoding is project called eMunch (Munch Museet, 2011), which is a repository of Edvard Munch's texts, postcards and notes.

6.4 IDA-presentation

Museums create a lot of different kind of compilation works - like video documentaries or slideshow presentations - in order to use them in exhibitions, different kind of events and seminars. A typical process involves searching images, videos and audio records from the museum's collection and then using some authoring program to construct the presentation. In this process media elements are copied to the presentation leaving no traces to the original element. Practically this means that information is packed for human understandable form into the presentation but at the same time this new context of the media elements is lost from the information retrieval point of view. By using the search engine terms, it can be said that the presentation is part of deep web.

The compilation work involves human authoring that creates new information and new connections between elements. Therefore any new compilation work itself creates a new context for media elements used. For example, when a presentation is made for museum's exhibition, it re-contextualises any media elements used and if that information could be "seen" by original elements that would enrich them. If presentation is made, for example with PowerPoint, a set of images are copied to the presentation from the collection management software and some texts are added (in worst case also the texts are copied from the same source) in order to contextualise the image.

From the information management's point of view compilation works made inside museums have several flaws. Compilation works duplicate information and materials and thus makes copied information outdated when original infor-

mation is modified. Secondly, if there is new information introduced in the presentation, that information is never feed into to the original information source. Third flaw of compilation works is that they require extra work even if they are merely duplicating information. In addition, the formats used are usually not suitable for long-term preservation.

There are two solutions for this problem: Either presentations can be indexed afterwards or a descriptive presentation format can be used in order to use original (ie. not copied) digital items in the presentation. In re-indexing the connections between original digital items and items used in the presentation are construed afterwards. In some cases this can be done automatically while other cases requires manual work. For example images in web pages can be automatically indexed if the file names of the images haven't changed from original. Links from the original file context (like collection management application) to the new web pages can be then created automatically. But if presentation is made by using descriptive meta format, media elements are not copied and the presentation can be queried for certain elements making elements aware about the new context.

IDA-presentation uses the latter solution: it is a XML-based presentation program. The file format of the IDA-presentation is heavily based on Synchronized Multimedia Integration Language (SMIL) (W3C, 2008). Using SMIL for cultural historical multimedia exhibitions is not new (Hong et al., 2005). However, the application is not a SMIL-viewer but the adaptation of the ideas presented in the SMIL. There are no slides in IDA-presentation. Instead the presentation is based on timeline and named regions, similarly to SMIL. With SEQ and PAR tags it is possible to control whether different parts of the content is presented sequentially or parallel. IDA-presentation adds two features to SMIL: support for 3D-models and IDA-framework queries.

Support for 3D-models in IDA-presentation is simple but suitable for presentation purposes. It mostly aimed for showing 3D-models from different angles and it allows exploded views for multi-part objects. The connection to IDA-framework makes it possible to fetch images or image information directly from the database. Therefore presentation can be made without duplication of data.

The file format of IDA-presentation is a descriptive format. A file defines how materials should be displayed without any platform of hardware dependent aspects like resolution of the screen. Latest development in HTML5 has shown that web browsers can be used for multimedia and even as a 3D platform. It would be quite possible to convert IDA-presentation to HTML5-presentations for example by using Popcorn.js. This conversion possibility highlights the idea of descriptive formats. These formats can be always reinterpreted for current platforms and hardware unlike, for example, HTML5 presentations with their tight integration to Javascript. XML-based format can be harvested for certain images or objects. This way it is possible to check if certain image or object is used in a presentation and this information can be attached to the original target.

In its current form IDA-presentation is not practical, since files must be hand written. But with a decent graphical user interface this would allow creation of platform independent, future-proof, time-based presentations for museum exhi-

bitions for example. The another thing is if the need for this kind of "documentation" presentation is seen by the museums. As one of the museum personnel said about the presentation in action: "It looks like a Powerpoint presentation". This illustrates the problem of "selling" the idea of sustainable information systems for the museums. If there is no any visible difference and if using the new system would mean abandoning the previous practises, then it is difficult convince users that this would be worth of trying.

Technically IDA-presentation is built in top of the OpenSceneGraph library (Osfield, 2008), which is an open source high performance 3D graphics toolkit and video support is implemented with FFMPEG.

6.5 IDA.video

Videos are good, fast and affordable method for documenting things in their context. Sounds and environment can be captured easily without any technical expertise. The problem is how to find separate parts, objects and concepts from the video afterwards without the need to watch entire video. The problem can be solved by linking certain moments in the video to a certain tags. Tags themselves can be anything from arbitrary names to categorised keywords or database items.

IDA.video is a simple, XML-based video tagging web application. The application demonstrates open design principles. Besides the open source code, openness means that all data (videos, tags and tag names) can be copied by anyone. The data is in XML format that can be accessed directly for information retrieval purposes. This allows users to copy or mirror data or they can integrate information they own applications or develop new user interfaces. IDA.video provides API for data manipulation. This means that it is possible to build alternative interfaces also for tagging, not just for viewing tagged videos.

Tags, attached tags and projects are stored in simple XML files. Because the data format is XML, it can be easily extended. Tags are just nodes in XML-file and the content of those tags can be defined by the users. They can include HTML or links to permanent URIs, for example. IDA.video is based on projects and tags associated to project. Every video is attached to a project and all videos in the project share same tags. This allows users to build tag structure that is specific for a certain perspective.

IDA.video is designed to be open. All data is kept in XML files that can be donwloaded and examined by anyone. Every project has its own tag vocabulary (tags.xml). One can study what tags have been created and how they have been applied to the videos. Only thing that obviously can not be copied is user's passwords. Passwords are placed inside a php-file that cannot be downloaded. This solution is a compromise between simple implementation and security.

The file structure of the IDA.video can be seen in Listing 6.4. The file files.xml lists all videos and defines in which project each video belongs. Every project directory contains tags.xml file that defines what tags are created. Tags_of_ - files

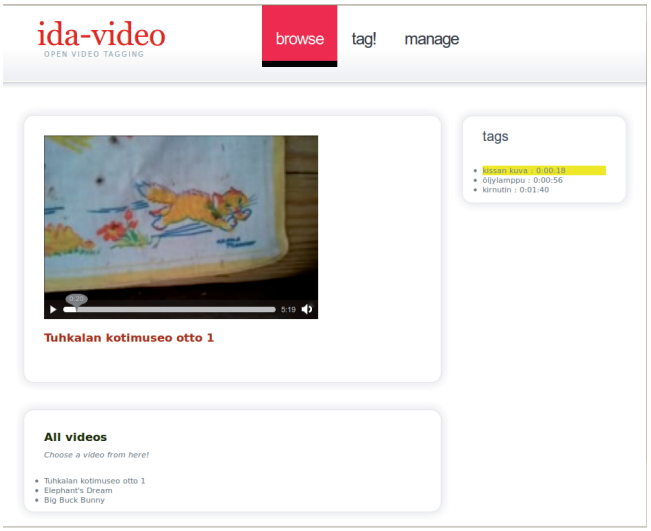


FIGURE 10 A screenshot of IDA.video. The tag ("image of the cat") has been clicked on the right panel and video has jumped accordingly showing the image of the cat on the video.

LISTING 6.4 File directories of IDA.video.

```
|-- files
|   |-- files.xml
|   |-- juva1.ogg
|-- projects
|   |-- homemuseum
|   |   | -- tags.xml
|   |   | -- tags_of_file_juva1_by_user_arihayri.xml
|   |   | -- tags_of_file_juva1_by_user_testuser.xml
|   |-- blender_films
|       .........
`-- users
```

lists attached tags for a certain file by a certain user in one file. The name of the file describes which file and which user is in question. The name of the file can be constructed by reading the content of the users directory and the content of files.xml.

Part of the openness is that there are some features that are expected to be implemented by the users. For example there is no search functionality of any kind. If user wants to implement a search for tag names, then the users must download tag files from each project and create a some kind of index of them in order to provide a search.

IDA.video is a prototype of applying open design principles to cultural heritage information systems. Although it is a small scale application and it was tested with only few videos, it still showed that these principles are functional. The application with its content can be easily moved to another PHP-enabled server and the content can be mirrored to another machine without permissions. The actual application logic can be re-implemented quite easily. Basically, the ap-

plication just need to keep list of videos and tags and offer ability to apply tags to the videos. All this information is kept in XML files that are publicly available. In addition, there must be a way to view videos with tags and a simple user rights system.

6.6 TUI.maps

TUI.maps was an experiment to build an tangible user interface for materials related to the campus area of Seminaarinäki. A tangible user interface means an interface that is used by manipulating (usually non-technical) objects, like pieces of wood or some kind of markers. The idea was to test the API provided by the IDA-framework in practise. TUI.maps is a concept of affordable and durable tangible user interface demo using IDA-QL as an information source. The main idea is to introduce a format that can be further utilised by "museum hackers" and to test the concept of hackable museum in practise. Demo is using devices that museums either already have (PC, data projector) or that can be purchased affordable (webcam). The source material is from the collection of the University Museum of Jyväskylä and the target area is the old campus area of the university.

Using tangible user interface with computer vision has some benefits in the museum environment. It does not need complicated devices, switches or wires. User interacts with "non-computer" items like paper or wooden blocks. Making these elements can even be part of the experience for example for children groups. This creates certain warmth compared for example of using sterile touch-screens. Tangible user-interface is also a very natural way to interact with the computer. One does not have to be afraid of accidentally touching the surface like in touch-based systems. System is also affordable, components can re-used in other projects and it is also durable since complex mechanical constructions are not needed.

TUI.maps is a concept of affordable and durable tangible user interface demo using IDA-QL as an information source. The main idea is to introduce a format that can be further utilised by "museum hackers" and to test the concept of hackable museum in practise. Demo is using devices that museums either already have (PC, data projector) or that can be purchased affordable (webcam). The source material is from the collection of the University Museum of Jyväskylä and the target area is the old campus area of the university.

The user interface of TUI.maps is any relatively flat surface with physical marker object or objects. The marker object has certain color or color combination, that makes it identifiable by means of computer vision. A webcam is pointing to the surface and the image is projected to the surface with data projector. A user studies the map by moving the marker to different places. The map can be a static, physical map or a projected image. When user moves the marker for example over a building, the system shows information and images about that target.

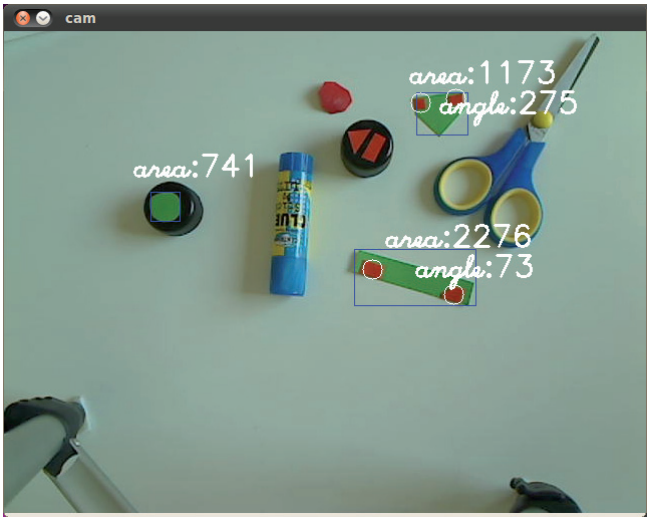


FIGURE 11 Markers among other objects have been detected by the tracking software. The green marker on the left has only position information but green markers with red dots have also orientation information.

There are two parts in the system; a tracker application and a display application. The tracker application uses simple algorithm to track markers. Markers are objects with certain color and size that can made of simply by cutting pieces of colored, non-glossy paper. In this case place markers of the buildings were placed by hand but the concept could be fully automated if buildings had geolocation, for example GPS-coordinates. When the map is placed in the same coordinate system, the place markers for the building could be created automatically. TUI.maps recognises two types of markers; position markers and orientation markers. Position marker is a marker that work as a cursor and by moving that, a user can study the area showed by a map. Orientation marker is similar to position marker but it has also orientation that can be tracked. Therefore it can be used as a virtual camera or as a generic pointing device when there is a need for aiming.

Top projecting introduces challenge to marker detection. Because images are projected on top, also markers receives the image. This makes detection harder since the colors and shapes are changing according the projected image. A simple solution for this problem is to design application so that there is no projection in top of the markers or that projected image does not affect the detection of markers.

A second version of the same interaction concept is a case where a full 3D-model of the area exists. Now also the marker's orientation is tracked and it serves as a virtual camera. The map serves as an orientation help for the user since user always knows where hers virtual viewer is located in the map. Similarly to the first demo, additional information can be provided when user places physical marker in the top of hot spots.

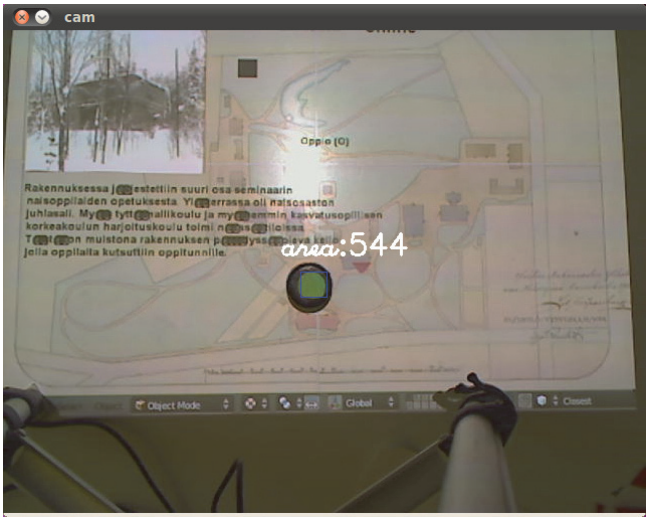


FIGURE 12 A View from the webcam: The webcam is placed in a tripod with the data projector. A green marker is inside the building hotspot and information about the building is shown. "Area" is debugging information and it means the size of the marker and black blobs in the text are caused by demo's insuffient support for scandinavian characters.

The actual application was built by using Blender and OpenCV computer vision library. Blender is an open-source 3D-modelling, game development and video editing software. Blender provides programmable real-time 3D environment by its game engine. Blender provides two ways for programming interactivity: Python programming language and logic bricks. Python provides a full programming environment that can be used for creating complex interactivity and games inside Blender. With logic bricks it easy to detect collisions between objects, for example. Together these features makes Blender efficient prototyping platform.

Blender is responsible of displaying content and fetching data from the database. Hotspots, areas that launches actions, are marked with 3D boxes in Blender scene. In this case, every boxes marks the location of a building and they every box has an ID that refers to the building's id in the database. When user places the marker in the top of building, the display application (Blender in this case) detects it and makes a IDA query for information and image about the building.

The tracking application was programmed with C++ using OpenCV computer vision library. Data transfer between Blender and the tracking application uses UDP protocol. This makes it possible to run the tracking software and the display software on different computers.

The TUI.maps demonstration proved that API developed for the IDA-framework was useful also for this kind of application development. All the materials could be fetched from the server to the application programmaticaly. Data and

materials are always up to date without any effort. However, from the long-term preservation point of view the demonstration has some weaknesses. First weakness is the use of API. Change in the API is likely to break the application. This is a generic problem with APIs since the change would break all applications using the same API. If the backend used XML files as a data storage, TUI.maps could have been implemented by directly accessing to XML-files. This could have been even simpler method since the application only fetched images of the buildings and some data like description, a building year and the name of the architect. By using XML files directly one avoids problems caused by changes in the API. The second problem is the platform itself. The Blender software is constantly changing and it is very unlikely that the demonstration will work in the future releases of Blender.

6.7 Demonstrations

Source materials of application demonstrations came mostly from the collections of University Museum of Jyväskylä. The Museum collects, preserves, makes research and exhibits material related to the past and the present of the University of Jyväskylä as well as the diversity of nature in Central Finland.

Maps

Two kinds of items were selected from the collections: maps and architectural drawings. There were about 200 maps and 100 architectural drawings in the museum's collection at the time of the experiment. A simple web application was created for browsing the maps. Maps can be browsed by the place, by the author or by the type of the map. The whole user interface was built with HTML and Javascript. More detailed explanation can be found from the conference paper called "A Template Based, Event-Centric Documentation Framework" later in this work.

Documentation of Seminarium building

Seminarium building was finished in 1883 and it is located in the campus area of University of Jyväskylä. The building was renovated in 2010 and image materials from different stages of the renovation was photographed by Riikka Mäkipelkola. About 350 images were selected and those images were inserted to IDA-framework. A special structure was needed in order to organise images. The building were split parts like rooms, hallways, stairs and windows (see Listing 6.5). The images were then attached to these parts so that they could be browsed more easily (see Figure 13).

Because database already included old images presenting different parts of the building, there was a need to separate renovation images from the old im-

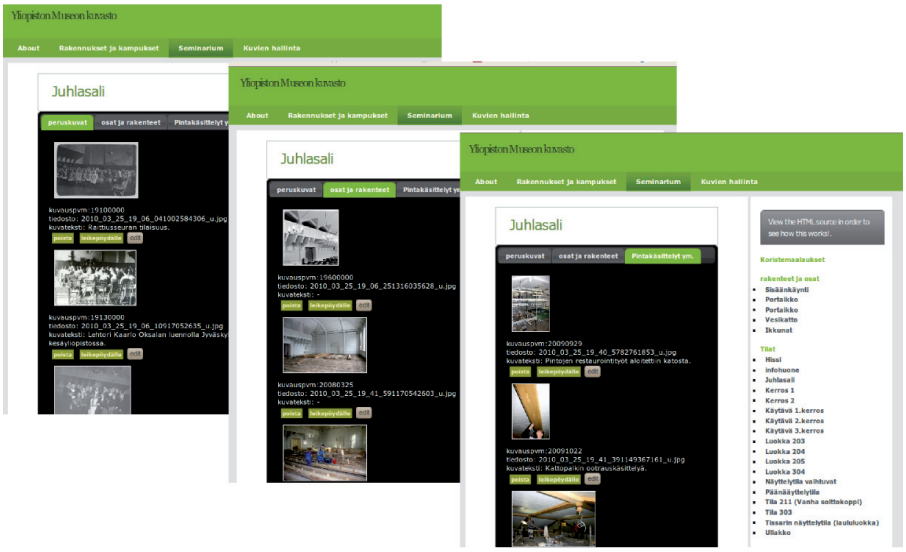


FIGURE 13 Three screenshots from the browsable user interface of the renovation images. From the left: showing base images of the room, structures of the room, showing coatings of the room.

ages. Obviously, the date that photograph was taken did separate images already but more semantic solution was needed. As a solution, the parts of the building were further divided - when needed - to structures like floor or ceiling. In addition, "coating" was added as a separate section for holding images that are representing painting or other surface operations. The same database was later used for creating a presentation to the opening ceremony of the Seminarium building.

LISTING 6.5 Semantic relations between building, rooms, structures and coatings. Simplified presentation based on CIDOC-CRM.

```
<Building name="Seminarium">
  <is_composed_of>
    <Room name="Main Hall">
      <is_composed_of>
        <Structure name="floor" />
      </is_composed_of>
      <bears_feature>
        <Coating name="decorative painting" />
      </bears_feature>
    </Room>
  </is_composed_of>
</Building>
```

Presentation of the renovation of the Seminarium building

IDA-presentation was used to construct a multimedia exhibition for the opening of the renovated Seminarium-building. Listing 6.6 shows an example of that presentation. The presentation shows renovation of the roof structure of the Seminarium building. Lines from 5 to 7 define regions where content is displayed.

The actual presentation starts from the line 12. The “seq” node starts a sequence that displays following nodes in order. The first node is an “img” node that defines the file name, duration and the display region. It also defines a callback, that fetches the meta information of the image from the IDA-framework. The image in question is a high detailed graph of the phases of the roof and it can not be read at average computer resolution at once. Therefore the image is shown in parts with panZoom property. A group of “animate” nodes defines what part of the image is visible and how long. In the line 19 there is a direct query that fetches all the images that present the roof (the roof is part of the building and its ID is 1519). Each image is shown 10 seconds except those that are defined in “replaceMedia” nodes. These nodes makes it possible to override the default behaviour and, for example, show individual files longer (lines 20 - 22) or hide some files (line 23).

The presentation is divided to several files. Each file presents a certain part of the building or a whole building. These files are combined as a single presentation by a master file which simply lists files in that order that they should be presented. This makes the presentation flexible and it makes easy to present only certain parts of the presentation.

LISTING 6.6 IDA-presentation file format.

```
1 <smil>
2  <head>
3   <layout>
4   <root-layout>
5    <regionImage id="full" left="0" top="0" width="100" height="100" z="-9" fit="meet"/
    >
6    <regionText id="IDA_imageNote" left="2" top="10" width="20" height="90" z="-3"
    backgroundColor="" fontSize="20"/>
7    <regionText id="title" left="2" top="2" width="100" height="100" z="-2" fontSize="
    30" >Vesikatto</regionText>
8   </root-layout>
9  </layout>
10  <transition id="fade" type="fade" dur="1s" />
11 </head>
12 <body>
13  <seq>
14   <img src="2010_03_28_18_49_071361764382_u.jpg" dur="114" region="full" callBack="
    IDA_all">
15    <animate attributeName="panZoom" begin="0" dur="4" from="0,0,100,100" to="
    0,0,100,100"/>
16    <animate attributeName="panZoom" begin="4" dur="1" from="0,0,100,100" to="
    0,55,45,45"/>
17    <animate attributeName="panZoom" begin="5" dur="19" from="0,55,45,45" to="
    0,55,45,45"/>
18   </img>
19   <query src="<search><CRM_Entity␣id='1519'/><result␣is_identified_by='name'></result
    ></search>"  region="full"  imageDur='10'>
20    <replaceMedia src="2010_03_28_18_49_191979369887_u.jpg" >
21     <img src="2010_03_28_18_49_191979369887_u.jpg" dur="30" region="full" />
22    </replaceMedia>
23    <replaceMedia src="2010_03_28_18_49_071361764382_u.jpg"></replaceMedia>
24   </query>
25  </seq>
26 </body>
27 </smil>
```

7 CONCLUSION

If we measure the importance of some institution to digital heritage by the amount of its digital objects, then heritage organisations are prominent among the central players. The problem is that many heritage organisations see digital heritage as an extension of tangible heritage that can be processed, controlled and authored in the same way as tangible heritage. But merely producing copyrighted digital surrogates and other materials is not enough any more. There are demands for sharing and enriching that material. The pressure is coming from two directions. First, when it comes to publicly funded museums, there are arguments that those materials should be public since public funds have been used for creating them. These arguments are quite similar to those in the Open Data movement. The second source of pressure is competitive information sources. These sources offer materials and information which not only overlap with what museums can supply; they often can be found more easily, and the quality of the material may actually be better than what museums could provide.

The Internet allows anyone to start their own virtual museum or blog about heritage issues without any permission from or quality control by heritage organisations. Search engines pick up all sources they can find regardless of organisational or educational status of the source's creators. These digital collections do not need original objects. They grow over institutional and national borders and overrun various control structures. If heritage institutions cannot - or are not willing to - answer to these demands, there is a risk that they will be bypassed in the digital world. Their role in education - and thus their role in society - is endangered. In the long term, this might have a negative affect also to the funding of these institutions.

Roadblocks

During my study, I was able to spot several structural issues that negatively affect the ability of the heritage organisations to participate in the creation of open digital heritage. These disadvantages can be dubbed an institutional burden of official heritage institutions. Some of the disadvantages are related to the prac-

tices of organisations, and some others to the official status of these organisations.

First, heritage institutions, like museums, are proud of their collections, which are the result of hard work and monetary investments. But in the networked, digital world a collection of physical assets can be a restrictive element. For example, a museum might have twenty percent of a certain artist's paintings in their collection. Publishing information about that missing eighty percent is not in the interest of the museum since they want to concentrate their efforts and resources on their own collection. For a person who is seeking information about the works of this specific artist this situation is not ideal. A tangible collection does not form a constraint to purely digital collections since materials can be collected from various sources. This renders museums second class citizens as information providers when it comes to information coverage. By default, as an information provider a museum is limited by its collections.

Secondly, legal issues, like copyright and privacy issues, are the area where museums are underdogs compared to non-institutional online sources. Copyright in Internet is far more flexible than what it is in official institutions since institutions must play by the rules. For example, a non-institutional online gallery may have copyrighted images from different sources while museums are not able to have those same images on display. The idea of fair use is commonly adopted in Internet communities. The legal status of fair use differs by country. For example, US legislation recognises fair use while Finnish legislation does not, which complicates the situation even more. In addition, privacy issues must be taken seriously. The names of persons and their actions are rapidly spread in the Internet when data is published. Thus, careful validation of the published data is required.

The third issue are the documentation practises of the museums and the tools that are used for documentation. Museum documentation practises have taken their form in the course of time. From the beginning of the history of museums, collection management and documentation have been done internally for internal purposes. The central point has been physical assets, and the target audience has been the museum expert. Suddenly these practises have become exposed to non-experts when museums have opened their online databases. Their cataloguing data has its value, but it might be too cursory or sparse for public use.

The tools museums use in cataloguing and managing collections are designed on these same principles. Being item-centric, the tools poorly support any different kind of information management. For example, there can be hundreds of images about the 400th anniversary of a town, with unique identifiers that can be referenced. A natural "umbrella" for those documents would be the anniversary of the town, but because anniversary is not an item it cannot be documented in this way. This kind of practice promotes documentation via objects instead of the actual targets, which leads to awkward documentation and problems in information retrieval, both internally and externally. Databases and database management systems are complex systems that are problematic not only for long-term preservation, but they also restrict the development of more flexible documenta-

tion models, such as spoken or free writing documentation. In addition, these systems set technical limitations to the openness and extensibility of the data.

The fourth problem of institutional digital heritage is the lack of resources. Funding of museums and other heritage institutions is not likely to increase according to new demands of openness and participation. If open digital heritage is wanted, then also performance indicators of museums must reflect the changed situation. From the website of the museum, open digital heritage materials can be spread around the Internet, where they can be re-used, combined and shared further. A website visitor counter won't provide a valid measurement of success any more, if it ever did. Performance indicators must take into account the effect that providing freely usable heritage materials entails.

The fifth aspect of institutional burden is the way these institutions position themselves regarding expertise and control. Heritage organisations' attitudes define how they see their role in the open information network. If a heritage organisation sees itself merely as an institution that collects, studies and exhibits cultural artefacts, then there is no room or need for delivering information outside the institution for free. Heritage institutions are expert organisations. The question is what that expertise means to the external world. For example, if heritage institutions see the Internet as a source of unreliable and misleading information that must be filtered out, then this means that expertise is a force of exclusion. Then control comes as an necessary part of expertise and this is the root of fears of museums.

Paths to More Open Digital Heritage

Digital assets produced by heritage organisations are needed in order to create open digital heritage. These assets are the basic building blocks that form the base of the new heritage. It is also important to realise that the question is not always about information or knowledge. It is just as important to expose people to heritage and heritage items in different contexts as it is to provide a single fully informed heritage channel. The contexts could be computer games, location-based heritage services, blog posts, mobile quiz applications, local virtual museums, online databases or school assignments. No single heritage organisation can provide all these contexts. Therefore, we need open access to allow anyone to create these contexts.

The institutional burden of the heritage organisations can be overcome. Although copyright issues are real and to be reckoned with, they can also be used as excuse to deny material from the public. Large portions of collection items have no copyright restrictions. For example, copyrights of collection item photographs usually belong to the museum. If items are not contemporary art, then there are usually no copyright restrictions for licensing them for free use even if other photographs in the collection could not be made available this way. Privacy issues must be taken into account also. However, documentation does not have to reveal the real identity of the people involved to outsiders. In many cases the actual names are irrelevant, as names can be substituted by random identifiers,

for example.

Further, heritage organisations must break the connection between expertise and control. The change from internal collection space to open information space in museums is possible only if museums see the value of open information that is not controlled by them. Expertise does not mean total control, it simply means valid knowledge that does not disappear if alternative views are presented. An expert organisation must be understood as a positive force with the ability to provide something valuable instead of rejecting everything external.

During my work, I have emphasised the importance of understanding the role of technology in digital heritage. I have shown that technical implementations affect the visibility of the museum materials to search engines, the flexibility of the documentation and long-term preservation. The resource and the documentation issues of the institutional burden are partly linked to these technical questions. A well-designed information system supports multiple documentation types, and openness of data does not require extra work, which, in turn, saves the resources of the organisation. Nevertheless, this requires changes in the information system design for heritage organisations.

A collection management software should do what its name implies: manage collections. A natural default unit for collection management is labelled tangible asset, and a software for this purpose can be implemented on that basis. However, if this fully item-centric approach is expanded beyond this particular task, then the only possible result is similarity to an item-centric system. Contextualising of cultural heritage objects cannot be done in such a system. Events, concepts, ideas and memories are not tangible objects. Therefore, ontological analysis is needed in order reveal all the target's relevant aspects that could be connected to each other. The result of the analysis can then be used in the process of creating a semantic framework.

So far ontological analysis is used mainly for information integration purposes. The CIDOC-CRM is a good example of this. What I am arguing is that ontological analysis should be extended to the documentation process itself. By revealing the essence of a documentable unit, it is possible to build a meaningful documentation platform. This requires a radical change in the tools used for documentation. They must support semantic documentation in a non-intrusive and flexible manner.

The design process of museum information system needs to be changed. Traditional database-driven systems are not the best solution when dealing with complexities of cultural heritage data. Systems must be open by default, and the role of tangible assets must be re-evaluated. Semantically precise and machine readable structures are needed, and these structures must be based on ontological analyses of collection items. Documentation of concepts, persons and events is equally important with the documentation of tangible assets. For small museums, these concepts, persons and events are mainly local entities with local nuances. Therefore, it is essential that these things can be defined by local actors. Local meanings can be built with local terms, not with global vocabulary lists or ontologies. Thus we can have open heritage systems that serve museums and

their needs without restricting the participation of the rest of the world.

One possible solution is to use file-based systems where the content is kept in open file formats. In an ideal case, data is semantically marked up, and semantic interference can be used for retrieval. But even if there were no semantic mark-up involved, the important thing with the file-based approach is that the contents of the system can be published as open data without any further procedures. Data can be mined and filtered by data hackers and data journalists if files are placed online. From the perspective of long-term preservation, file-based systems have two benefits. The content of a file-based system can be easily mirrored to another server. Basically, one line of code is needed in order to keep an up-to-date copy in a local machine. The second benefit is a direct consequence of the first one: content can be stored for long-term preservation since it is plain textual data and other data formats used are open data formats. In addition, this approach is suitable also for less resourceful institutions since no extra effort is needed for opening the data.

Information systems in the age of the new heritage need to be open, simple and extensible. The requirement for simplicity in these systems is crucial, especially for small organisations. If the system used is simple, less expertise is required to modify or even rebuild the system, as the time passes. This means that the modifications are cheaper and the risk of outdated system is smaller. The complexity needed can be built by linking several simple systems together via universal resource identifiers. When these identifiers are systematically used in the internal system, they can be easily used also externally. This extensibility is crucial so that new forms of documentation and participation can be implemented and tested also outside of the organisation. One lesson learnt from the software developed during this work was that there is no user interface design that would suit to all users. Therefore, it is important to allow user interfaces to be created independently by the users, who know what suits best for them.

However, despite all the complexity of technical issues and their pertinence for the openness of the digital heritage, sharing in the name of Free Culture might simply mean that images provided by a heritage organisation are licensed under a free licence. No APIs, ontologies, XML or semantics are necessarily needed, just simple web pages with image descriptions and invitation to participate by re-using and re-defining visual heritage.

Finally

My definition of open digital heritage is wide. For me, open digital heritage means shared responsibility for our digital cultural heritage. I see digital heritage as something that should not be owned by museums, libraries or archives. I would go even as far as to say that digital heritage is one form of free speech. Defining, redefining, contextualising, re-contextualising and representing our cultural objects is our fundamental right. But this does not mean neglecting the value of professional heritage keepers. Experts are still needed. Professionally run museums are still needed. Tangible objects are fragile, and they are best

kept safe in the hands of trained professionals without any intervention from outsiders. But digital heritage is different. Digital objects do not break when handled without any expertise. They do not wear out even if handled by millions. Their meaning might change, but everyone is entitled to transform the meaning as they please. Experts and heritage institutions should not prevent outsiders from being part of the creation of digital heritage.

Open digital heritage is not a technical issue, nor is it a non-technical issue. In order to develop open digital heritage, a holistic view is needed. The nature of the open Internet, the framework of museums, design principles of information systems, rights of people to define their own heritage and the need to preserve digital information to future generations are all aspects that must be taken into account when creating new heritage. Non-technical problems cannot be solved by technical means, but, at the same time, technical aspects must be known by those who professionally work with digital heritage.

I defined two views for open digital heritage: Free Culture view and Open Data view. By liberating cultural historical materials for human beings and opening cultural historical data for machines, heritage organisations can support the democratisation process of digital heritage. New heritage is creating its forms and its needs. We do not know them all yet. All we know is that these new forms require access to digital heritage and permission to make that material personally meaningful.

8 EPILOGUE

While searching online cases for my work I stumbled upon a list of 3D objects made for a computer game called Operation Highjump. Operation Highjump is a crowdsourced game loosely based on a movie called Iron Sky. Iron Sky, in turn, is a crowdsourced movie project that tells about Nazis' revenge in the year 2018 after they had been hiding in the dark side of the moon.

One of the props was a frame carrying Arnold Böcklin's Isle of Death (Figure 31). I am sure that many heritage professionals might think that the use of Böcklin's work in a possibly humorous (possibly, since the game has not been released yet) Nazi-genre computer game is misuse of great art. But is there any problem if a player is exposed to the artwork of Böcklin while gaming? Maybe the strong, mystical and dark aesthetics of Isle of Death raises a desire to experience more these kinds of works. It might be the first time when the player actually pays attention to a painting. This is an ultimate escape of art from exhibition galleries to virtual worlds. Again, how bad can that be?

FIGURE 14 Arnold Böcklin's (1827 - 1901) Isle of Death on the surface of a 3D-model for the Operation Highjump game

YHTEENVETO (FINNISH SUMMARY)

Digitoitu kulttuuriperintö muodostaa erittäin tärkeän - joskaan ei ainoan - osan digitaalisesta kulttuuriperinnöstä. Digitoitujen aineistojen rooli on palvella esteettömyyttä mutta ennen kaikkea tarjota lähtökohtia tulkinnoille ja keskusteluille. Ihmisten mahdollisuus osallistua digitaalisen kulttuuriperinnön määrittelyyn on riippuvainen siitä miten helppoa näiden digitaalisten objektien löytäminen ja käyttäminen on. Digitaalinen kulttuuriperintö ei kuitenkaan ole välttämättä avointa. Se voi olla vaikeaa löytää esimerkiksi siksi että hakukoneet eivät pysty indeksoimaan suljetuiksi suunniteltuja online-tietokantoja tai löytyvä materiaali saattaa olla tehty käyttökelvottomaksi vesileimoilla, joiden tarkoitus on ainoastaan tuhota digitaalinen toisinto. Kysymys kuuluukin: Haluavatko kulttuuriperintöorganisaatiot olla mukana luomassa avointa digiperintöä vai haluavatko ne pitää kiinni erityisasemastaan kulttuuriperinnön määrittelijöinä.

Vaikka muistiorganisaatioilla on hyvin vankka asema myös digitaalisen kulttuuriperinnön tuottajina, ne eivät kuitenkaan ole ainoita materiaalin tuottajia. Niiden kilpailijoiksi on ilmaantunut tiedon tuottajia, jotka eivät ole sidoksissa kulttuuriperintöaineistojen virallisiin tuottajiin. Nämä uudet tiedonlähteet eivät ole vain päällekkäisiä traditionaalisiin muistiorganisaatioihin verrattuna vaan ne saattavat tarjota kattavampaa ja relevantimpaa tietoa. Hakukoneet indeksoivat materiaalia katsomatta sitä kuka materiaalin tuottanut. Tuloksena on kenttä jos institutionaalinen digitaaliperintö ja ei-institutionaalinen perintö tapaavat ja törmäävät.

Internet on kiehtova ja jopa anarkistinen tietoverkko, jonka käytännöt ovat venyttäneet esimerkiksi tekijänoikeuden ja yksityisyydensuojan käsitteitä. Museomaailmalla puolestaan on oma, pitkäikäinen traditio aineistoihin suhtautumisessa. Kun museot ja muut muistiorganisaation julkaisevat materiaalejaan Internetissä, syntyy mielenkiintoinen ristiriita kahden hyvin erilaisen kulttuurin välillä. Työssäni tarkastelen tätä törmäyspintaa ja digitaalista kulttuuriperintöä vakiintuneiden muistiorganisaatioiden ulkopuolelta. Tutkin esimerkkitapausten kautta kuinka organisaatioiden materiaali näkyy jokapäiväisessä Internet-verkossa. Hakukoneet ja niiden tapa järjestää aineistoa muodostavat tärkeän osa jokapäiväisestä Internetin käytöstä. Kuinka institutionaalinen digiperintö näkyy hakukoneissa? Entä kuinka avoin Internet luo kilpailua vakiintuneita kulttuuriperinnön toimijoita kohtaan?

Toinen tutkimuskysymykseni koskee teknologian osuutta digitaalisen kulttuuriperinnön avoimuudessa. Teknologia on aina läsnä puhuttaessa digitaalisesta kulttuuriperinnöstä ja siksi teknologian ymmärtäminen on tärkeää esimerkiksi pohdittaessa digitaalisten aineistojen avoimuutta tai pitkäaikaissäilytystä. Museoiden kokoelmanhallinnalla on pitkä perinteet ja myös niiden tietojärjestelmät perustuvat pitkälti näihin perinteisiin. Järjestelmät on enimmäkseen suunniteltu museoiden sisäiseen käyttöön ja ne myös ohjaavat dokumentointia. Tietojen julkaiseminen ulkopuoliseen käyttöön on sangen uusi vaatimus ja sen suhteen erilaiset ratkaisut Järjestelmä vaikuttaa siihen millaista dokumentointia

museo pystyy tekemään. Tekniikka voi mahdollistaa joitain asioita mutta se voi myös vaikeuttaa tai tehdä mahdottomaksi toisia. Siksi informaatiojärjestelmien merkitys on merkittävä dokumentoinnin kannalta.

On mahdollista löytää useita seikkoja jotka hankaloittavat avoimen digitaalisen kulttuuriperinnön syntymistä. Osaa näistä esteistä voidaan nimittää institutionaaliseksi taakaksi sillä ne liittyvät muistiorganisaatioiden viralliseen asemaan ja niiden käytäntöihin. Osa tekijöistä on taas on puhtaasti teknisiä. Ensinnäkin, fyysinen kokoelma on museon ylpeyden aihe johon on panostettu aikaa ja rahaa. Mutta kokoelma voi olla taakka Internet-maailmassa. Jos museo omistaa kaksikymmentä prosenttia taiteilijan teoksista, se ei välttämättä ole halukas tarjoamaan tietoa puuttuvasta kahdeksastakymmenestä prosentista. Tiedonhakijan kannalta tilanne ei tietenkään ole ideaalinen. Tällaista rajoitetta ei ole digitaalisilla kokoelmilla. Museo on siis oletuksena rajoittunut tiedontarjoaja ja tämän rajoittuneisuuden takana on fyysinen kokoelma.

Toinen alue, jossa museot ja muut viralliset kulttuuriperintöorganisaatiot ovat altavastaajina ei-institutionaalisiin Internet-resursseihin verrattuna on tekijänoikeudet. Monet Internetin käytännöt ovat muokanneet tekijänoikeuden käsitettä joustavampaan suuntaan. Niin sanottu "fair use" -periaate on internetyhteisöissä sangen yleisesti hyväksytty tapa käsitellä tekijänoikeudenalaista materiaalia. Tämä joustavuus ei kuitenkaan ulotu muistiorganisaatioihin, sillä niiden on toimittava kulloisenkin lainsäädännön mukaan. Kolmas avoimen digitaaliperinnön ongelmista on museoissa on niiden dokumentointikäytännöt ja työkalut, joita tässä työssä käytetään. Dokumentointikäytännöillä on pitkä historia ja lähes yhtä pitkään dokumentoinnin kohde on ollut fyysinen esine ja dokumentoinnin kohdeyleisö on ollut kyseisen alan asiantuntija. Tämä tekee aineistoista vaikeasti lähestyttäviä ulkopuolisille.

Neljäs hyvin oleellinen tekijä on museoiden ja muiden muistiorganisaatioiden resurssit. Vaikka vaatimukset aineistojen avaamisesta ja osallistumisesta verkkoyhteiskuntaan kasvavat, ei nähtävissä kuitenkaan ole resurssien kasvamista samassa suhteessa. Viides tekijä institutionaalisessa taakassa on asiantuntijuus ja kuinka organisaatio liittää sen kontrolliin. Jos museo näkee Internetin vain epäluotettavan ja väärän tiedon lähteenä verrattuna omaan asiantuntijuuteensa, niin silloin asiantuntijuuden ylläpitäminen vaatii rinnalleen kontrollin jolla estetään tämän epäluotettavan tiedon julkaiseminen. Tämän vuoksi aineistoja ei haluta antaa yleiseen käyttöön, mikä puolestaan estää ulkopuolisia tahoja osallistumasta digitaalisen kulttuuriperinnön rakentamiseen.

Institutionaalinen taakka on olemassa mutta se ei estä museoita ja muita vastaavia organisaatioita osallistumasta avoimen digitaalisen kulttuuriperinnön luomiseen, jos ne haluavat siihen osallistua. On tärkeää että kulttuuriperinnön parissa toimivat tahot ymmärtävät että asiantuntijuus ei edellytä kontrollia. Asiantuntijuus on nähtävä positiivisena voimana, jolla on jotain tarjottavaa, eikä voimana jolla suljetaan muut näkemykset ulkopuolelle.

Tekijänoikeudet ja yksityisyydensuoja on luonnollisesti otettava huomioon aineistoja jaettaessa. Toisaalta tekijänoikeudet antavat kätevän tekosyyn olla julkaisematta aineistoja vaikka mitään tekijänoikeudellista estettä ei olisikaan.

Suuri osa aineistoista on kuitenkin tekijänoikeusvapaita. Esimerkiksi museon kokoelmaesineistään ottamat valokuvat voidaan yleensä julkaista vapaan lisenssin alla, jos kuvat eivät esitä tekijänoikeuden alaisia teoksia. Yksityisyydensuojan suhteen oleellista on se, että useinkaan yksityisyyden suojan alainen tieto, kuten yksittäisten henkiöiden nimet, ei ole oleellista materiaalin julkaisemista ja jatkokäyttöä ajatellen. Lahjoittajatieto tai esineen käyttöhistoria ei välttämättä tarvitse henkilöiden nimiä ollakseen hyödyllistä informaatiota ulkopuolisille. Nimet voidaan korvata esimerkiksi satunnaisilla merkkijonoilla julkisessa tiedossa, jos tämä on otettu huomioon tietojärjestelmän suunnittelussa.

Olen työssäni halunnut korostaa teknologian ymmärtämisen merkitystä digitaalisen kulttuuriperinnön kanssa toimiessa. Teknologia on aina läsnä puhuttaessa digitaalisesta kulttuuriperinnöstä ja se vaikuttaa digitaalisen informaation avoimuuteen, laatuun ja pitkäaikaissäilytykseen. Tämän vuoksi tutkimukseni sisältää käytännön ohjelmistokokeiluja, joiden avulla olen selvittänyt millaisin suunnitteluperiaattein avointa digitaalista kulttuuriperintöä voidaan edesauttaa. Perustamalla järjestelmän hajautettuun arkkitehtuuriin monoliittisten järjestelmien sijaan ja käyttämällä yksinkertaisia, tekstitiedostoihin perustuvia järjestelmiä informaation tallettamiseen, voidaan samalla parantaa sekä digitaaliperinnön avoimuutta että sen pitkäaikaissäilytystä. Näin data voidaan avata avoimen datan periaatteita noudattaen kun samalla aineistojen pitkäaikaissäilytys helpottuu. Tämä malli sopisi erityisesti pienille toimijoille.

Teknologia liittyy kahteen institutionaalisen taakan kysymykseen eli resursseihin ja työkaluihin, joilla digitaalista aineistoa tuotetaan. Hyvin suunnitellussa järjestelmässä aineistojen avoimuus ei aiheuta lisätyötä, mikä puolestaan säästää resursseja. Kokoelmanhallinnan perusyksikkönä on ollut fyysinen objekti ja tätä kautta myös museoiden tietojärjestelmistä on tullut fyysisiin objekteihin perustuvia. Tämä on toimiva malli pelkän kokoelmanhallinnan osalta mutta ei kokoelman kontekstualisoinnin osalta. Tämän ongelman ratkaisemiseksi tarvitaan ontologista analyysia parempien järjestelmien aikaansaamiseksi. Tähän mennessä ontologinen analyysi on nähty lähinnä työkaluna aineistojen yhtenäistämisessä, kuten esimerkiksi CIDOC-CRM:n tapauksessa. Mutta analyysi täytyy ulottaa itse dokumentointiin jotta voidaan selvittää millaisia suhteita esineiden, tapahtumien, ihmisten ja ilmiöiden välillä on mahdollista dokumentoida.

REFERENCES

Abid, A. 2007. Safeguarding our digital heritage: a new preservation paradigm. In Y. D. Lusenet & V. Wintermans (Eds.) Preserving the digital heritage Principles and policies. Selected papers., Vol. 31. Amsterdam: European Commission on Preservation and Access, 7–14.

Akmon, D. 2010. Only with your permission: how rights holders respond (or don't respond) to requests to display archival materials online. Archival Science 10 (1), 45–64. doi:http://dx.doi.org/10.1007/s10502-010-9116-z.

Anon. 2011. Kana nimeltä Ross 508 - syötkö vielä kanaa? [discussion forum]. ⟨URL:http://www.pakkotoisto.com/kotitalous/97840-kana-nimeltae-ross-508-syoetkoe-vielae-kanaa/⟩. accessed 2012 Feb.

Arnold, D. & Geser, G. 2008. Framework. Budapest: ARCHAEOLINGUA, 244. ⟨URL:http://www.epoch-net.org/index.php?option=com_content&task=view&id=200&Itemid=306⟩.

Auletta, K. 2010. Googled. The End of the World As We Know It. The Random House, 418.

Bekiari, C., Constantopoulos, P. & Doerr, M. 2005. Information design for cultural documentation. In In Proceedings of the 9th DELOS Network of Excellence thematic workshop: Digital Repositories: Interoperability and Common Services. ⟨URL:http://www.ics.forth.gr/isl/publications/paperlink/Informationdesign.pdf⟩.

Bitzer, J., Schrettl, W. & Schröder, P. J. H. 2005. Intrinsic Motivation in Open Source Software Development. EconWPA. Development and Comp Systems. ⟨URL:http://ideas.repec.org/p/wpa/wuwpdc/0505007.html⟩.

Bognár, K., Karjalainen, M., Saraniva, A., Szász, B. & Unzeitig, M. 2006. Cultural Heritage on the Semantic Web - the Museum24 project. ⟨URL:http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/articles/Szasz-museum24Paris.pdf⟩.

Bogost, I. 2009. A Television Simulator: CRT Emulation for the Atari VCS. ⟨URL:http://www.bogost.com/games/a_television_simulator.shtml⟩. accessed 17.11.2011.

Cameron, F. 2007. Digital Knowledgescapes: Cultural, Theoretical, Practical, and Usage Issues Facing Museum Collection Databases in a Digital Epoch. In F. Cameron & S. Kenderdine (Eds.) Theorizing Digital Cultural Heritage A criticial discource, 165–191.

Cameron, F. 2008. Politics of Heritage Authorship: The case of digital heritage collections. In Y. E. Kalay, T. Kvan & J. Affleck (Eds.) New Heritage. New Media and Cultural Heritage. Routledge, 170–184.

Castells, M. 2007. Communication , Power and Counter-power in the Network Society. International Journal of Communication 1 (June 2006), 238–266.

Chesal, R. 2012. Loosen up copyright law, says Dutch government. ⟨URL:http://www.rnw.nl/english/article/loosen-copyright-law-says-dutch-government⟩. accessed 3.5.2012.

Chiles, A. 2008. Lecturer bans students from using Google and Wikipedia. ⟨URL:http://www.theargus.co.uk/news/1961862.lecturer_bans_students_from_using_google_and_wikipedia/⟩. accessed 10 August 2011.

Covey, T. 2005. Acquiring Copyright Permission to Digitize and Provide Open Access to Books. Library Administration And Management October, 63. ⟨URL:http://www.clir.org/pubs/reports/pub134/pub134col.pdf⟩.

Crofts, E. N., Doerr, M., Gill, T., Stead, S. & Stiff, M. 2006. Definition of the CIDOC Conceptual Reference Model. ⟨URL:http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.1.pdf⟩. accessed May 2008.

Dixon, L. 1985. THE 1960 BOWDEN SPACELANDER! ⟨URL:http://nbhaa.com/indexBowden.html⟩. accessed 2012 Sep.

Doering, Z. D. 1999. Strangers, Guests, or Clients? Visitor Experiences in Museums. Curator: The Museum Journal 42 (2), 74–87. doi:10.1111/j.2151-6952.1999.tb01132.x.

Doerr, M. & Iorizzo, D. 2008. The dream of a global knowledge network - A new approach. Journal on Computing and Cultural Heritage 1 (1), 1–23. doi:10.1145/1367080.1367085. ⟨URL:http://portal.acm.org/citation.cfm?doid=1367080.1367085⟩.

Doerr, M. & Kritsotaki, A. 2006. Documenting Events in Metadata. In The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST. ⟨URL:http://cidoc.ics.forth.gr/docs/fin-paper.pdf⟩.

Donovan, K. 1997. The Best of Intentions : Public Access , the Web & the Evolution of Museum Automation. In Museums and the Web 1997 Conference. ⟨URL:http://www.archimuse.com/mw97/speak/donovan.htm⟩.

Ellis, M. 2009. If you love something, set it free. ⟨URL:http://www.slideshare.net/dmje/if-you-love-something-set-it-free-presentation⟩. accessed 2012 Sep.

Ellis, M. 2011. Managing and Growing a Cultural Heritage Web Presence. A strategic guide. London: Facet Publishing, 224.

Eschenfelder, K. R. 2009. Controlling Access to and Use of Online Cultural Collections. IMLS, 89. ⟨URL:http://digital.library.wisc.edu/1793/38251⟩.

Eschenfelder, R. & Caswell, M. 2010. Digital Cultural Collections in an Age of Reuse and Remixes. ⟨URL:http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3060/2640⟩. accessed 11 August 2011.

Europeana 2011a. Europeana search. ⟨URL:http://www.europeana.eu/portal/search.html?query=james+bond⟩. accessed August 2011.

Europeana 2011b. The Europeana Licensing Framework, 20. ⟨URL:https://version1.europeana.eu/c/document_library/get_file?uuid=3f32e3b6-b69b-40b5-92a6-32654571087f&groupId=10602⟩.

Google 2011. Webmaster Tools help Frames. ⟨URL:http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=34445⟩. accessed 12 August 2011.

Gruber, T. R. 1993. A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications. Knowledge Creation Diffusion Utilization 5 (April), 199–220.

Häyrinen, A. 2008. A template based, event -centric documentation framework. ⟨URL:http://cidoc.mediahost.org/content/archive/cidoc2008/Documents/papers/drfile.2008-06-26.pdf⟩.

Häyrinen, A. 2009. IDA-framework development site. ⟨URL:www.opendimension.org/ida⟩.

Häyrinen, A. 2011. Towards Semantic Modelling of Cultural Historical Data. In N. Heimbürger, A., Kiyoki, Y., Tokuda, T., Jaakkola, H., Yoshida (Ed.) Information Modelling and Knowledge Bases XXII. IOS Press, 312 – 320.

Häyrinen, A. 2012. Source code of the dissertation "Open sourcing Digital Heritage : digital surrogates, museums and knowledge management in age of open networks". ⟨URL:http://urn.fi/URN:NBN:fi:jyu-201210042581⟩.

Hall, S. 2002. Whose Heritage? Un-settling the 'Heritage', Re-imagining the Post-Nation. In R. Araeen, Z. Sardar & S. Cubitt (Eds.) The Third text reader: on art, culture, and theory. Continuum, 72–84.

Hands, J. 2011. @ is for Activism Dissent, Resistance and Rebellion in a Digital Culture. London: Pluto, 210.

Harley, D. 2007. Why Study Users? An environmental scan of use and users of digital resources in humanities and social sciences undergraduate education. ⟨URL:http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1423/1341⟩. accessed 11 August 2011.

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. Farnham, UK: Ashgate, 35–52. ⟨URL:http://sebastianheath.com/publications⟩.

Himanen, P. 2001. The hacker ethic and the spirit of the the information age. New York: Random House, 232.

Hong, J.-S., Chen, B.-H., Hung, S.-H. & Hsiang, J. 2005. Toward an integrated digital museum system - the Chi Nan experiences. International Journal on Digital Libraries 5 (3), 231–251. doi:10.1007/s00799-005-0116-1. ⟨URL:http://www.springerlink.com/index/10.1007/s00799-005-0116-1⟩.

Huynh, D. F. s. a. Timeline software. ⟨URL:http://simile.mit.edu/timeline/⟩.

ISO 2006. Conceptual Reference Model in ISO. ⟨URL:http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue%_detail.htm?csnumber=34424⟩. accessed May 2008.

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

Janée, G., Mathena, J. & Frew, J. 2008. A data model and architecture for long-term preservation. In Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries - JCDL '08. New York: ACM Press, 134. doi:10.1145/1378889.1378912.

Jones-Garmil, K. 1997. Laying the Foundation: Three Decades of Computer Technology in the Museum. In K. Jones-Garmil (Ed.) The Wired Museum (2nd edition). American Association of Museums, 34–62.

Lerner, J. & Tirole, J. 2001. The open source movement: Key research questions. European Economic Review 45 (4-6), 819–826. doi:10.1016/S0014-2921(01)00124-6.

Lessig, L. 2004. Free culture : how big media uses technology and the law to lock down culture and control creativity. New York: The Penguin Press, 345. ⟨URL:http://www.free-culture.cc/freecontent/⟩.

Lindner, J. 2006. Digitization Reconsidered. ⟨URL:http://www.media-matters.net/docs/JimLindnerArticles/DigitizationReconsidered.pdf⟩.

Lonkila, H. 2006. 3D-bridge final report. ⟨URL:http://www.arthis.jyu.fi/bridge/tir.html⟩.

Lunt, B. M., Hyatt, D., Zhang, F., Saini, G. & Linford, M. R. 2008. Preserving Our Digital Heritage. In 2008 Digest of Technical Papers - International Conference on Consumer Electronics. Ieee, 1–2. doi:10.1109/ICCE.2008.4587848. ⟨URL:http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4587848⟩.

MacArthur, M. 2007. Can Museums Allow Online Users to Become Participants? In H. Din & P. Hecht (Eds.) The Digital Museum A Thinkg Guide. American Association of Museums, 57–66.

Manovich, L. 2008. Softaware takes command. ⟨URL:http://lab.softwarestudies.com/2008/11/softbook.html⟩.

Marty, P. F. 2008a. Changing Needs and Expectations. In P. F. Marty & K. B. Jones (Eds.) Museum Informatics People, Information and Technology in Museums. New York: Routledge, 182–185.

Marty, P. F. 2008b. Information Representation. In P. F. Marty & K. B. Jones (Eds.) Museum Informatics People, Information and Technology in Museums. New York: Routledge, 29–34.

Mayer-Schönberger, V. 2009. Delete The Virtue of Forgetting in the Digital Age. Princeton University Press, 235.

McLean, K. 2004. Museum Exhibitions and the Dynamics of Dialogue. In G. Anderson (Ed.) Reinventing the Museum. AltaMira Press, 193–211.

Mudge, M., Ashley, M. & Schroer, C. 2007. A digital future for cultural heritage. In Symposium A Quarterly Journal In Modern Foreign Literatures, 01–06. ⟨URL:www.isprs.org/proceedings/XXXVI/5-C53/papers/FP104.pdf⟩.

Nikšić, H. s. a. Wget. ⟨URL:http://www.gnu.org/s/wget/⟩.

Onnela, T. 2011. Digitoidut kuvakokoelmat verkossa, kenen käytössä ja millä ehdoilla? ⟨URL:http://agricola.utu.fi/keskustelu/viewtopic.php?f=10&t=3460&p=17572#p17572⟩. accessed August 2011.

OpenExhibit 2010. OpenExhibit Surveys. ⟨URL:http://openexhibits.org/surveys⟩. accessed Sep 2012.

Osfield, R. 2008. OpenSceneGraph. ⟨URL:http://www.openscenegraph.org/⟩.

Parry, R. D. 2009. Digital Heritage and the rise of theory in museum computing. Museum Management And Curatorship 20 (4), 333–348. ⟨URL:http://dx.doi.org/10.1016/j.musmancur.2005.06.003⟩.

Peacock, D., Ellis, D. & Doolan, J. 2004. Searching For Meaning: Not Just Records. In D. Bearman & J. Trant (Eds.) Museums and the Web 2004. Arlington: Archives & Museum Informatics. ⟨URL:http://www.archimuse.com/mw2004/papers/peacock/peacock.html⟩.

Peacock, D. 2008. The Information Revolution in Museums. In P. F. Marty & K. B. Jones (Eds.) Museum Informatics People, Information and Technology in Museums. New York: Routledge, 59–76.

Pitkänen, M. 2007. Reikäkortti - kuka vielä muistaa? ⟨URL:http://www.tietovaki.fi/lehti/arkisto/Tietovaki_2007_1.pdf⟩.

Polso 2009. Alvar Aallon urkufasadit. ⟨URL:http://p2.foorumi.info/urkufoorumi/viewtopic.php?t=163⟩. accessed June 2011.

Porter, J. s. a. American Bison at the Castle, 1887 - 1889. ⟨URL:http://newsdesk.si.edu/snapshot/american-bison-castle⟩. accessed August 2011.

Raymond, E. 1999. The cathedral and the bazaar. Knowledge, Technology & Policy 12 (3), 23–49. doi:10.1007/s12130-999-1026-0. ⟨URL:http://www.springerlink.com/index/10.1007/s12130-999-1026-0⟩.

Raymond, E. 2001. How To Become A Hacker. ⟨URL:http://www.catb.org/esr/faqs/hacker-howto.html⟩. accessed September 2012.

Raymond, E. 2004. The Jargon File. ⟨URL:http://www.catb.org/jargon/⟩. accessed September 2012.

Rieger, O. Y. 2009. Search engine use behavior of students and faculty: User perceptions and implications for future research. ⟨URL:http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2716/2385⟩. accessed Aug 2011.

Rogozanu, C. 2009. Relax, mai sint si carti misto pe lumea asta. Ideile si evolutionismul. ⟨URL:http://voxpublica.realitatea.net/popcult/relax-mai-sint-si-carti-misto-pe-lumea-asta-ideile-si-evolutionismul-20773.html⟩. accessed August 2011.

Ruotsalo, T. & Hyvönen, E. 2007. An event-based approach for semantic metadata interoperability. In Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference. Berlin, Heidelberg: Springer-Verlag. ISWC'07/ASWC'07, 409–422. ⟨URL:http://dl.acm.org/citation.cfm?id=1785162.1785193⟩.

Savov, V. 2010. Kinect finally fulfills its Minority Report destiny. ⟨URL:http://www.engadget.com/2010/12/09/kinect-finally-fulfills-its-minority-report-destiny-video/⟩. accessed Feb 2011.

Shirky, C. 2005. Ontology is Overrated: Categories, Links, and Tags. ⟨URL:http://www.shirky.com/writings/ontology_overrated.html⟩. accessed Sep 2011.

Silfverberg, A. 2012. Mitä ajattelin tänään: Frankensteinin perhetila. ⟨URL:http://nyt.fi/20120120-mitae-ajattelin-taenaeaen-frankensteinin-perhetila/⟩. accessed January 2012.

Simons, N. 2008. The Future of Authority: Platform Power. ⟨URL:http://museumtwo.blogspot.com/2008/10/future-of-authority-platform-power.html⟩.

Sinclair, P., Lewis, P., Martinez, K., Addis, M. & Prideaux, D. 2006. Semantic web integration of cultural heritage sources. In WWW '06: Proceedings of the 15th international conference on World Wide Web. New York, NY, USA: ACM Press, 1047–1048. ⟨URL:http://doi.acm.org/10.1145/1135777.1136008⟩.

Sola, T. 1997. Essays on museums and their theory : towards the cybernetic museum. Helsinki: Finnish Museums Association, 293.

Sutela, H. 2011. Syö mieluummin isoa eläintä. ⟨URL:http://www.hs.fi/ruoka/artikkeli/Sy%C3%B6+mieluummin+isoa+el%C3%A4int%C3%A4/1135263006557⟩. accessed Feb 2012.

Tammisto, S. & Häyrinen, A. 2006. Modelling Seminaarinmäki buildings with Inkscape and Blender. ⟨URL:http://opendimension.org/ida5D/inkscape.php⟩. accessed December 2011.

Terras, M. 2010. Digital curiosities: resource creation via amateur digitization. Literary and Linguistic Computing 25 (4), 425–438. doi:10.1093/llc/fqq019. ⟨URL:http://llc.oxfordjournals.org/cgi/doi/10.1093/llc/fqq019⟩.

Unsworth, B. Y. J. M. 2004a. The Next Wave : Liberation Technology. ⟨URL:http://www3.isrl.illinois.edu/~unsworth/liberation.html⟩.

Unsworth, B. Y. J. M. 2004b. The Value of Digitization for Libraries and Humanities Scholarship. ⟨URL:http://www3.isrl.illinois.edu/~unsworth/newberry.04.html⟩. accessed 25.2.2011.

Weber, S. 2000. The Political Economy of Open Source Software. ⟨URL:http://brie.berkeley.edu/econ/publications/wp/wp140.pdf⟩.

Weber, S. 2004. Source. Harvard University Press. ⟨URL:http://site.ebrary.com/lib/jyvaskyla/Doc?id=10328787&ppg=102⟩.

Weil, S. E. 1999. From being about something to being for somebody. Daedalus (Boston) 128 (3), 229–258. ⟨URL:http://www.jstor.org/stable/20027573⟩.

Willis, N. 2007. Lessons learned from open source Xara's failure. ⟨URL:http://archive09.linux.com/feature/119790⟩. accessed Sep 2012.

Wilson, D. 2010. Microsoft furious at $2,000 bounty for open source Kinect drivers. ⟨URL:http://news.techeye.net/software/microsoft-furious-at-2000-bounty-for-open-source-kinect-drivers⟩. accessed 20.2.2011.

Zorich, D. M. 1997. Beyond Bitslag*: Integrating Museum Resources on the Internet. In K. Jones-Garmil (Ed.) The Wired Museum. American Association of Museums, 171–201.

Adafruit Industries 2010a. Open Kinect driver(s) released. ⟨URL:http://www.adafruit.com/blog/2010/11/10/we-have-a-winner-open-kinect-drivers-released-winner-will-use-3k-for-more-hacking-plus-an-additional-2k-goes-to-the-eff/⟩.

Adafruit Industries 2010b. The Open Kinect project. ⟨URL:http://www.adafruit.com/blog/2010/11/04/the-open-kinect-project-the-ok-prize-get-1000-bounty-for-kinect-for-xbox-360-open-source-drivers⟩. accessed 20.2.2011.

Apache Software Foundation 2011. How the ASF works. ⟨URL:http://www.apache.org/foundation/how-it-works.html#meritocracy⟩. accessed December 2011.

Brooklyn Museum s. a. Bowden Spacelander. ⟨URL:http://www.brooklynmuseum.org/opencollection/objects/2587/Spacelander_Bicycle⟩.

European Comission 2000. The Welfare of Chickens Kept for Meat Production (Broilers). ⟨URL:http://ec.europa.eu/food/fs/sc/scah/out39_en.pdf⟩.

Free Software Foundation 2007a. GNU General Public License, version 3. ⟨URL:http://www.gnu.org/copyleft/gpl.html⟩. accessed December 2011.

Free Software Foundation 2007b. Why Open Source misses the point of Free Software. ⟨URL:http://www.gnu.org/philosophy/open-source-misses-the-point.html⟩. accessed December 2011.

Gramps project 2011. Gramps. ⟨URL:http://gramps-project.org/⟩.

Kansallinen Digitaalinen Kirjasto 2011. Hallinnolliset ja rakenteelliset metatiedot ja aineiston paketointi. ⟨URL:http://www.kdk.fi/images/stories/tiedostot/kdk_metatiedot_ja_aineiston_paketointi.pdf⟩.

Munch Museet 2011. eMunch. ⟨URL:http://www.emunch.no:8080/cocoon/emunch/⟩.

Mygreenweek s. a. CRAZY WEDNESDAY: WONDERFUL NEW ELECTRIC BIKE!!! ⟨URL:http://senseup.wordpress.com/?s=bowden⟩. accessed Feb 2011.

National Board of Antiquities s. a. Cultural Environmental Portal. ⟨URL:http://kulttuuriymparisto.nba.fi/netsovellus/rekisteriportaali/portti/default.aspx⟩. accessed August 2011.

Nordiska Museet 2011. James Bond's toy gun. ⟨URL:http://digitaltmuseum.se/things/thing/S-NM/NM.0279468A-B⟩. accessed August 2011.

Oikeutta Eläimille 2012. Oikeutta eläimille -yhdistys teki tutkintapyynnön Atrian mainonnasta. ⟨URL:http://www.oikeuttaelaimille.net/www2/lang-fi/ajankohtaista/tiedotteet/341-oikeutta-elaeimille-yhdistys-teki-tutkintapyynnoen-atrian-mainonnasta⟩. accessed February 2012.

Online Computer Library Center (Ed.) 2010. Perceptions of Libraries, 2010: Context and Community. OCLC. ⟨URL:http://www.oclc.org/reports/2010perceptions.htm⟩.

Open Data Foundation 2011. The Open Data Foundation. ⟨URL:http://www.opendatafoundation.org/⟩. accessed December 2011.

Open Knowledge Foundation 2011. What Is Open? ⟨URL:http://opendatamanual.org/what-is-open-data/what-is-open-data.html⟩. accessed December 2011.

Ross Breeders Ltd 2000. Broilerien hoitokäsikirja, 112. ⟨URL:http://www.virtuaali.info/2/pdf/ross_broilerienhoitokasikirja2000.pdf⟩.

Smithsonian Institute 2009. Smithsonian Institute: Web and New Media Strategy. Smithsonian Institute, 26. ⟨URL:http://www.si.edu/content/pdf/about/web-new-media-strategy_v1.0.pdf⟩.

Tekijänoikeusneuvosto 2011. LAUSUNTO 2011:3. ⟨URL:http://www.minedu.fi/export/sites/default/OPM/Tekijaenoikeus/tekijaenoikeusneuvosto/tekijaenoikeusneuvoston_lausunnot/2011/TN_2011-3.pdf⟩.

U.S Copyright Office 2009. Factsheet on fair use of copyrighted works. ⟨URL:http://www.copyright.gov/fls/fl102.html⟩. accessed 18.11.2011.

UNESCO 2010. Concept of Digital Heritage. ⟨URL:http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/preservation-of-documentary-heritage/digital-heritage/concept-of-digital-heritage/⟩. accessed 5.11.2010.

W3C 1999. Frames in HTML documents. ⟨URL:http://www.w3.org/TR/html4/present/frames.html⟩. accessed August 2011.

W3C 2008. Synchronized Multimedia Integration Language. ⟨URL:http://www.w3.org/TR/2008/REC-SMIL3-20081201/⟩.

Wikipedia s. a. Google bomb. ⟨URL:http://en.wikipedia.org/wiki/Google_bomb⟩. accessed December 2011.

Wikipedia s. a. Open-source hardware. ⟨URL:http://en.wikipedia.org/wiki/Open-source_hardware⟩. accessed 16.11.2011.

Wikipedia s. a. Vuoksenniskan kirkko. ⟨URL:http://fi.wikipedia.org/wiki/Kolmen_Ristin_kirkko⟩. accessed Feb 2011.

APPENDIX A KYSELY

Hei, teen väitöskirjaa digitaalisen kulttuuriperinnön ja yhteisöllisen tiedontuottamisen mahdollisuuksista. Olen kiinnostunut niistä esteistä, joita museoilla voi olla aineistojen vapaaseen käyttöön ja uudelleentulkintaan liittyen digitaalisissa medioissa. Toisin sanoen etsin museoiden mukavuusalueen rajoja digitaalisten aineistojen suhteen.

Sarka-museo on maatalouden museon ja maatalouteen liittyy nykyaikana erottamattomasti kysymykset ympäristökuormituksesta ja eläinten oikeuksista.

Yksi kohteistani on Tipitii-näyttely, joka teillä oli esillä viime vuonna. Näyttely oli käsittääkseni Suomen Broileryhdistyksen tuotantoa: http://www.siipi.net/index.php?option=com_content&view=article&id=80:tipitii-suomen-broileri tuotanto-50-vuotta.

Internetistä on helppo löytää näyttelyn aineistolle hyvinkin vastakkaista kuva-aineistoa broilerituotannosta ja toisaalta paljon kritiikkiä broilerituotannon etiikka ja käytäntöjä kohtaan. Tämä ristiriita ei kuitenkaan ulottunut millään tavoin näyttelyyn. Siksi haluaisin kysyä museonne kantaa seuraaviin kysymyksiin:

kysymys 1: Koetteko että teillä on velvollisuus julkista rahoitusta saavana museona tuoda esiin erilaisia näkökulmia myös silloin kun eturistiriita näyttelyn tekijän/rahoittajan kanssa on olemassa?

****

vastaus 1: Totta kai koemme velvollisuudeksemme saada näyttelyissämme julki mahdollisimman monipuolinen, todellinen ja edustava kuva kuvattavasta asiasta. Toisaalta viittauksesi internetin tietoihin ja kuviin on sekin vain yksi puoli: pääsääntöisesti liiankin kriittinen ja johonkin enemmän yksittäistapaukseen nojaava. Meillä ei ole eturistiriitaa näyttelyn tekijän tai rahoittajan kanssa, koska pyrimme toimimaan rehellisesti ja puhtaaseen objektiiviseen tietoon nojaten. Me emme pohjaa näyttelyitämme niitä tehdessämme internetin kritiikittömiin ja sattumanvaraisiin tietoihin vaan oikeaan ja todelliseen sekä kattavaan tietomateriaaliin.

Internetin sijasta kehottaisin Sinua käymään ihan analogisesti jollakin broileritilalla katsomassa, miten kaikki oikeasti toimii. Internetistä ei saa kuvaa sen paremmin broilerituotannosta kuin meidän näyttelymmekään kokonaisuudesta.

****

kysymys 2: Koetteko että olisitte halutessanne voineet tuoda esiin eläinoikeusnäkökohdan (tai eläinoikeuksien "nousun" merkityksen broilerituotantoon) Tipitii-näyttelyssä?

vastaus 2: Eläinten hyvä ja asiallinen kohtelu on aivan ilman muuta suomalaisen kotieläintalouden kulmakiviä, jos kohta varmasti tästä löytyy laajalta tuottajakentältä poikkeuksiakin (nämä päätyvät internettiin). Tavallisen tuottajan päämääränä on päästä mahdollisimman hyvään tuotantoon sekä määrällisesti ja laadullisesti, ja nämä päämäärät eivät varmasti toteudu, jos eläimiä pidetään huonosti. Siksi kysymyksesi on hieman provosoiva. Totta kai tällainen

eläinoikeusnäkökulma on olemassa, mutta näemme, että se on aivan luonnollinen lähtökohta kaikessa karjataloudessa. Siksi sen korostaminen tuotantoa käsittävässä suhteellisen suppea-alaisessa näyttelyssä ei voi olla sen keskeisin näkökulma. Broilerituotantoa on harjoitettu Suomessa 50 vuotta ja tänä aikana tapahtunut kehitys ja muutos on se keskeinen seikka, joka tässä näyttelyssä piti olla pääroolissa.

****

kysymys 3: Oletteko museona ajatelleet tehdä eläinten oikeuksiin maataloudessa tai maatalouden ympäristökuormitukseen liittyvää näyttelyä?

vastaus 3: Eläinoikeudet sinänsä on varsin filosofinen kysymys, jonka rinnastaminen esimerkiksi ihmisoikeuksiin ei voi olla aivan analoginen. Karjataloutta harjoitetaan esimerkiksi broilerituotannossa yksinomaan lihantuotantoa silmällä pitäen. Jos eläinoikeudeksi määritellään hyvä ja täysipainoinen elämä, ei teurastukseen päättyvä tuotanto voi sitä luonnollisesti koskaan olla. Toisaalta hyvän elämän jälkeinen ja luonnolliseen kuolemaan päättyvä elämä ei ole esimerkiksi broilerituotannossa ollenkaan mahdollinenkaan. On tietenkin kokonaan filosofinen kysymys, halutaanko tuottaa lihaa syötäväksi vai ei. Jos lihaa tuotetaan syötäväksi, tuotannossa joudutaan väistämättä ainakin sen viimevaiheissa sortamaan eläinoikeuksia, mikäli ne halutaan ymmärtää niiden kaikista jyrkimmässä muodossa.

Emme toistaiseksi ole ajatelleet tehdä näyttelyä eläinoikeuksista maataloudesta, mutta voimme ilman muuta sellaistakin ajatella. Meillä ei ole olemassa tähän mitään arvoasetelmaa tai aatteellista estettä. Paljon tärkeämpi kysymys ympäristömme osalta on minun mielestäni maatalouden ympäristökuormitus ja sen vähentäminen. Tätähän koetetaan koko ajan edistää kaikkien osapuolten toimesta niin meillä kuin muuallakin, sillä ympäristölle aiheutettu kuorma on käytännössä hukkaan menevä menoerä kasvinviljelyn kannalta katsottuna. Tätä koskeva näyttely voisi olla hyvinkin hyödyllinen ja pidän mahdollisena sellaisen tekemistä jossakin tulevaisuudessa. Museomme on kuitenkin pieni ja valmistamme keskimäärin yhden uuden näyttelyn vuodessa. Tämän vuoksi näyttelypolitiikkamme on väistämättä pitkäjänteistä. Näytettävää olisi kovin paljon ja mahdollisuudet niin kovin pienet.

****

kysymys 4: Koetteko että ympäristökysymykset tai eläinoikeuskysymykset voisivat olla kiusallisia kohdeyleisöllenne?

vastaus 4: Emme missään tapauksessa koe, että nämä kysymykset olisivat mitenkään kiusallisia kohdeyleisöllemme. Suomessa on harjoitettu maataloutta runsaat 3000 vuotta ja se on tästä ajasta pitkälle yli 2900 vuotta ollut kansamme pääelinkeino. Museomme ei missään tapauksessa ole olemassa vain tuottajayhteisöä varten, vaan tehtävämme on kertoa kaikille suomalaisille, mistä leipä ja maito tulevat ruokapöytiimme ja miten niiden tuotantotavat ovat muuttuneet ajassa. Nämä asiat tulisi jokaisen suomalaisen ymmärtää, jotta osaisimme arvostaa tätä elinkeinoa. Jos maataloutta ei Suomessa harjoitettaisi, maa pystyisi keräilyn ja pyynnin avulla elättämään kaikkiaan 15 000 henkeä. Maatalous oli se voima, joka mahdollisti Suomen väestön kasvun noin kolmeen miljoonaan. Se

mahdollisti suomalaisen kansan synnyn ja omaleimaisen kulttuurin muodostumisen.

Kohdeyleisömme on siis koko kansa ja ulkomaiset asiasta kiinnostuneet näiden lisäksi. Siksi kysymyksesi on provosoiva ja asiaa ymmärtämätön.

****

kysymys 5: Onko teillä ollut tai oletteko harkinneet turkistarhaukseen liittyvää näyttelyä? Koska aihe epäilemättä herättää vastakkainasetteluja, niin

vastaus 5: Vastauksessa 3 esitin jo näyttelypolitiikkaamme kohdistuvia rajoitteita. Mikään ei estäisi meitä tällaista näyttelyä tekemästä, mutta sellainen ei tällä hetkellä ole näyttelyohjelmassamme.

****

kysymys 6: Oletteko harkinneet sosiaalisen median käyttöä osana toimintaanne?

vastaus 6: Käytämme sosiaalista mediaa jo nyt toiminnassamme ja olemme käyttäneet sitä jo varsin pitkään. Meillä on olemassa oma Facebook-profiili ja sillä lukuisa määrä ystäviä. Harjoitamme sen kautta tiedottamista ja yhteydenpitoa ympäristöömme.

****

Oletteko saaneet yhteydenottoja eläinoikeus- tai ympäristöjärjestöiltä?

7) Aivan joku yksittäinen yhteydenotto on muistaakseni ollut, mutta se oli hyvin yleinen eikä se oikeastaan kritisoinut toimintaamme.

****

Olisin kiitollinen vastauksista. Käytän mahdollisia vastauksianne osana tutkimukseni aineistoa.

All of the papers are conference papers. The annual CIDOC conferences (Committee for Documentation of the International Council of Museums) were my primary arena for publishing my work. Three of my papers were presented in these conferences. The abstracts of the papers were reviewed by the program committee and based on acceptance, the full papers were provided. There were no paper publications made from CIDOC conferences.

ORIGINAL PAPERS

PI

MEMORIES IN THE WAREHOUSE. DEVELOPING A MULTIPERSPECTIVE INFORMATION SYSTEM FOR CULTURE HISTORICAL USE

by

Ari Häyrinen 2006

CIDOC 2006, Göteborg, Sweden

Memories in the warehouse. Developing a multiperspective information system for culture historical use

The focus of museums has shifted from objects to phenomena. This change is also reflected in the information systems that are used in museums. Traditional systems have serious difficulties when it comes to storing the new approach. This means that, in current systems, a museum object has only one voice: the official one.

In order to provide museum objects with multiple voices, a new role must be given to the museum's information system. Instead of seeing it as a static archive, the information system is considered as a tool for the interpretation process, in which the insertion of the object is the starting point. In the 3D-Bridge project Jyväskylä University Museum has started to develop an application for multiperspective information handling. Currently it is used for virtual reconstruction of the campus area.

IDA makes a strict separation between targets and documents. A document is any digital file (like image or 3D-model) that is uploaded to the system. After the upload the document is attached to a target or targets, since documents cannot exist in the system without being linked to a target. The target can be a collection item or an event or any other object stored in the database. Any target can also be a document. For example, photograph is a collection item (target) but at the same time, it can be linked as a document to some other target.

Arguments are IDA's way of handling multiple views to the core data. Arguments are html-documents with text and links to the database material. Arguments belong to the project and are therefore organized by projects. Arguments are constructed with a built-in WYSIWYG editor. It is possible to use any item from the database in an argument through IDA's link board. When an item is picked from the link board to the argument, it is actually linked to it. Links are always bidirectional, so every item also "knows" when it has been used in the argument, and the link back to the argument is shown among the item's other information.

IDA also supports 3D models as a one type of document. However, in order to make the actual real-time presentation, the separate 3D client is beeing developed. IDA3D is based on an open-source rendering library called OpenSceneGraph. IDA3D creates a virtual world based on the models and information provided by IDA. Models are placed on layers. There could be separate layers, for example, ground, vegetation, wooden houses or buildings that were planned but never actually built. This makes it easy to compare and, if necessary, to isolate different parts of reconstruction with just one button press. One important aspect is that because the models are attached to the targets, 3D models preserve their context and meta-information in IDA3D.

In recent years the focus of museums has shifted from object-centricity to a more anthropocentric view. However, from the information system design's point of view, it is very tempting to outline the museum as a collection of objects. That is, to see the museum as a place that stores objects and holds some infor-

mation attached to them, like a warehouse. This allows the system designer to take advantage of traditional database design patterns that are suitable for stock accounting. This approach, however, is doomed to fail.

What makes this issue more complex is the fact that museum is a collection. However, storing objects is just one part of museums' nature. Museums (like libraries and archives) are social memory organizations; they are a collection of research material and a collection of interpretations. They even are a collection of empty spots, things that are not in the collection. The meaning of a museum's collection is something that is defined by a viewer's point of view. For an animal activist, a collection of fur farming tools and leather production is a collection of equipment used for mistreating animals. On the contrary, from the research's point of view, those tools may be seen as a timeline that shows the technical development of fur production. For a retired farmer those tools are full of memories from the past. Still, it is the same collection, only the interpretation varies.

The change in the focus of museums is also reflected in the information systems that are used in museums. Traditional systems perform poorly when it comes to storing alternative interpretations in the system. Overall, there are a very limited number of views to the material available. There may be information about donation, restoration, ownership, subject index and some notations about context. All these views are predefined by the system designer and adding a new view is usually impossible without a redesign of the whole database. There is no room for the farmer's memories or for the animal activist's view. This means that, in current systems, a museum object has only one voice: the official one.

In order to provide museum objects with multiple voices, a new role is given to the museum's information system. Instead of a static archive, the information system must be seen as a dynamic network. In archive mentality an object is stored in the system and the process inside the archive stops there: the object is ready when it is saved. However, if the information system is considered as a tool for the interpretation process, the insertion of the object is the starting point of the process. The dialog can continue through the system, and the dialog process remains visible. Some parts of the net remain unchangeable, but new parts keep coming in, and those new parts change the way older parts are seen. This approach is very useful for research purposes, for example, for culture historical reconstruction because it gives a tool for the documentation of the research process. By following the dialog, the roots of the interpretation can be seen.

In the EU-2000 project called 3D-Bridge we at Jyväskylä University Museum started to develop an application for culture historical reconstruction. As part of the project, we wanted to reconstruct our university's campus area in its 1920's appearance as a social milieu and an architectural entity. In order to do that, we needed a tool that would allow different types of views to the core material and could be used for work with interactive 3D-models. Two applications were developed for this purpose: a web-based information system called IDA (Ideal Documentation Archive) and IDA3D for virtual reality aspects.

IDA combines database structure and hyperlinking in order to produce a multiperspective information system that allows an unlimited number of differ-

ent types of views to the protected core data.

IDA makes a strict separation between targets and documents. A document is any digital file (like image or 3D-model) that is uploaded to the system. After the upload the document is attached to a target or targets, since documents cannot exist in the system without being linked to a target. The target can be a collection item or an event or any other object stored in the database. Any target can also be a document. For example, photograph is a collection item (target) but at the same time, it can be linked as a document to some other target.

IDA's way of handling multiple views to the core data are arguments. Arguments are html-documents with text and links to the database material. Arguments belong to the project and are therefore organized by projects. Arguments are constructed with a built-in WYSIWYG editor. It is possible to use any item from the database in an argument through IDA's link board. When an item is picked from the link board to the argument, it is actually linked to it. By following the link in the argument, it is always possible to find the original data. Links in IDA are always bidirectional, so every item also "knows" when it has been used in the argument, and the link back to the argument is shown among the item's other information. For example, a transcription of a farmer's interview could be inserted in the database as an argument. Then, images of the tools that are mentioned could be linked to the text. As a result, there would be an illustrated transcription that is easy to follow even for a person that does not recognize the names of the tools.

From the accessibility perspective IDA's arguments are one way to approach material instead of traditional queries. Queries can be problematic for the general public because their efficient usage requires knowledge of databases in general, and some information about the content of the database is needed. Projects create a new analysis of the material, but at the same time, a new user interface is created. One might say that every project in IDA is a mini-exhibition, which has a unique perspective to the material.

Unlike in many systems, images in IDA are treated in a non-atomic way. Usually images in databases are treated as atomic objects, i.e. something that cannot be divided in smaller parts. However, it is very common that the content of the image is not one, unified object. Therefore, the system must provide tools for handling parts of an image. IDA supports linked cropping of images. Linked cropping creates a new image document that can be used separately from the original image. But unlike with traditional cropping, the connection between images still remains. It means that the cropped image always knows where it was cropped from, and the original image is aware of all images that were cropped from it.

Some interesting possibilities appear when the non-atomic treatment of images is combined with the argumentation structure. For example, if a study is made on how the Virgin Mary is painted in altar paintings, every appearance of the Virgin Mary could be cropped from the altar painting images. Then, the researcher could write notes about these images using them as illustration in the arguments. Now, there would be a new perspective to the altar paintings based

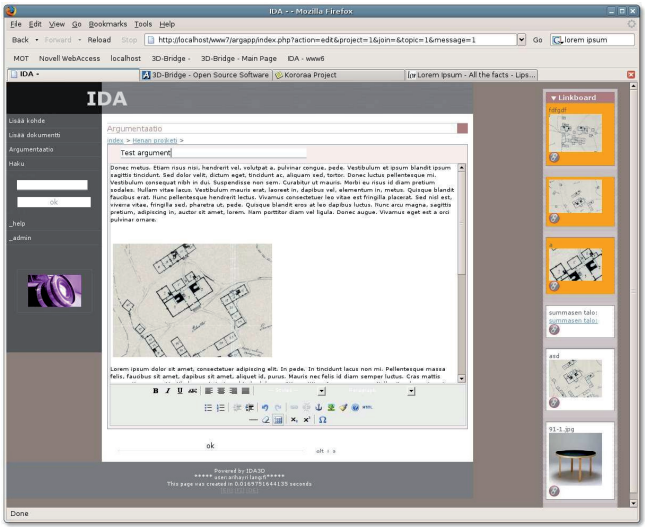


FIGURE 15 The argument editor. The linkboard and its content is on the right.

on only one of their aspects. Or in the case of art historical vocabulary, the examples could be cropped from actual targets in IDA. The Doric pillar example could be cropped from the image of the Doric temple. However, now there would also be a link from the Doric temple to the vocabulary and, therefore, the content of the Doric temple would be enriched.

The other part of IDA is its ability to store 3D models. In IDA, a 3D model is just one type of document. However, in order to make the actual real-time presentation, the separate 3D client was developed. IDA3D is based on an open-source rendering library called OpenSceneGraph. IDA3D creates a virtual world based on the models and information provided by IDA. Models are placed on layers. There could be separate layers, for example, ground, vegetation, wooden houses or buildings that were planned but never actually built. This makes it easy to compare and, if necessary, to isolate different parts of reconstruction with just one button press. One important aspect is that because the models are attached to the targets, 3D models preserve their context and meta-information in IDA3D.

The IDA system will be accessible for everyone interested in it. Both programs are released under a free GPL license and currently they are under heavy development. More information can be found on the project's website: http://www.arthis.jyu.fi/ida/index.php.

About Jyväskylä University Museum

Jyväskylä University Museum The Seminaarinmäki Hill in Jyväskylä has housed museum activities for already a hundred years. In fact, the beginning of the museum's collections can be dated to the 1860s. The Jyväskylä Teacher Training College, the first Finnish training institute for elementary school teachers, was established in 1863, and started to collect didactic specimens. When the Col-

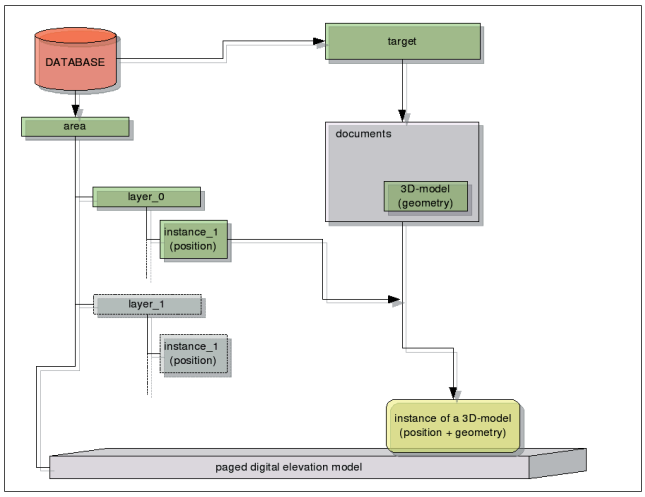


FIGURE 16 Structure of the real-time visualisation system.

lege of Education became a University of Jyväskylä, the museum also received the status of university museum. In 1980, it was divided into two separate departments: the Section of Cultural History and the Section of Natural Sciences. Over the decades, the Jyväskylä University Museum has evolved from an ethnological collection maintained by training college students into a museum which presents the history, functioning, and the general cultural scene of the academic community. The museum functions as a training post for museology; a teaching museum where students can receive practical training in museum work. website: http://www.jyu.fi/tdk/museo/museum.html .

PII

A TEMPLATE BASED, EVENT-CENTRIC DOCUMENTATION FRAMEWORK

by

Ari Häyrinen 2008

CIDOC 2008, Athens, Greece

A Template Based, Event-Centric Documentation Framework

By using explicit events in documentation and ontology based on CIDOC-CRM, it is possible to built a very flexible documentation application. However, CIDOC-CRM itself is too abstract to be used as an information model and it was never meant to be used as such. Therefore there must be a more concrete layer of semantics top of CIDOC-CRM that defines the structure of an individual record. This was accomplished by using documentation templates, which are used to describe special features of different types of documentation targets. The developed framework offers easy-to-use, open-source framework for small organisations and makes adoption of semantic-aware systems possible for wider audience by hiding the complexity of semantic ontologies.

Introduction

Museums are experts on documentation and good documentation applications should respect that knowledge. If documentation application has a very rigid structure, it may prevent museum to make documentation of their area of interest in an appropriate way. Therefore documentation applications should be flexible in order to fit different kind of purposes even in the same domain. However, this flexibility decreases interoperability between different documentation data since the data structure of applications differs between them. This problem can be solved by mapping data to a core ontology like CIDOC-CRM (Crofts et al., 2006). Nevertheless, if the data structure is highly dynamic in order to reach required flexibility, the mapping process comes much more difficult. But if there is a shared seman- tic core and mechanism to define a record type specific semantics top of it, then it is possible to achieve semantic interoperability between different data sets without losing flexibility.

Properties and classes defined in CIDOC-CRM provide a good model for event-centric documentation, even though it is claimed that core ontologies like CIDOC-CRM are not needed in order to achieve semantic interoperability (Ruotsalo and Hyvönen, 2007). CIDOC-CRM's strength is that event-centric approach is built in it and it has predefined structure for handling them through properties. CIDOC-CRM has been created on empirical bases from real-world datasets which reflects the special needs of cultural historical field and CRM also has also status of being an ISO-standard (ISO, 2006). However, very little attention is given to possibilities and possible problems event-centric approach creates for making new documentation. Most of the research efforts and application development around CIDOC-CRM, and semantic interoperability of cultural historical data in general, are related to problems concerning of mapping already existing material to ontologies and theoretical is- sues behind that process. An event-centric ontology is a working solution for information integration but this approach can also be applied for documentation work itself. By using core ontology that is based on CIDOC-CRM and documentation templates, it is possible to create a very flexible documentation tool that still produces semantically interoperability data.

Explicit events in cultural historical documentation

An event-centric documentation model means that events are explicitly present to the person who is making a documentation. Explicit event-centric documentation approach is used for example in genealogy application called GRAMPS (Gramps project, 2011). Birth, death and marriage are some of the event types that can be added to a person record and events can have their own documentation. GRAMPS can also produce different kind of reports of person. Museo24 -project (Bognár et al., 2006) uses CIDOC-CRM as core ontology and events are explicitly present in annotation tool but that requires at least some knowledge about CIDOC-CRM. The event-centric approach has several advantages compared to traditional, non-event-based ap- proaches. First, human actions are the cultural context of museum items and that must be expressed somehow in documentation. Events provides semantically meaningful way of describing links between things and actions of human beings (Doerr and Kritsotaki, 2006). Secondly, the event-centric model allows very flexible structure for an individual record. Events form a history of an item and new events can be added at any time. In practise this means, that documentation can be very detailed or just in general level and that the level of details can be decided by the person making the documentation. Therefore documentation can be constructed based on qualities of the documentation target, instead of rigid structure that documentation application provides. Using explicit events also simplifies data structure design because events have a common structure: someone did something somewhere in certain time period. With this formulation it is possible to represent for example a creation of an artwork, a construction of a building or having a scientific seminar. The third, and maybe the largest benefit of all, is that using explicit events in documentation makes events themselves documentable units. This allows more detailed, and less item-centric, documen- tation since it is possible to represent the history of the target with different kind of events. For example, it is possible to document design process of buildings or restoration of paintings. Events split documentation into smaller units therefore making documentation semantically more precise and more accessible. The fourth benefit of using explicit events in documentation is the fact that because events are basically time periods, the processing of time is relatively simple. One can cre- ate cross-sections of time easily and see what events of which target occurred during that time pe- riod. Using explicit events also allows very easy creation of timeline presentations, like Timeline (Huynh, s. a.). In order to realise practical application using event-centric documentation model, a dynamic data structure is needed and the user-interface design must reflect that structure. A highest level of documentation flexibility would be achieved if there were just properties that user would pick up as they are needed. This is not a very practical solution because that would require a lot of knowledge about underlying ontologies and properties. Therefore a mechanism is needed for hiding the complexity of ontologies and defining what properties are typical for different kind of records.

Documentation template

The idea of a semantic documentation template is to provide a record type specific documentation frame that can be defined by organisation responsible of documentation. The template defines a typical case for a record including default properties, required events, possible events and possible parts. More precisely, documentation template maps a part of thesauri to a partial domain ontology that is build on top of CIDOC-CRM. It is a reversed mapping that is done before actual data input but with additional information of how documentation should be made. It also serves a guide for a good documentation of a specific type of record. Here is a very simple example template for a person:

```
<class id="E21" title="Person">
  <property id="P131F" title="is_identified_by" required="1">
    <table id="personname"/>
    <order>1</order>
    <action>input</action>
  </property>
  <property id="P98B" title="was_born" required="0">
    <class id="E67" title="Birth" />
    <order>2</order>
    <action>default_event</action>
  </property>
  <property id="P100B" title="died_in">
    <class id="E69" title="Death" />
    <order>2</order>
    <action>possible_event</action>
  </property>
</class>
```

A person must be identified by name and every person must have a birth event, since this is re- quired for person's existence. Death of a person is defined as a possible event. Birth itself has its own template:

```
<class id="E67" title="Birth">
  <property id="P96F" title="by_mother">
    <class id="E21" title="Person" />
    <order>2</order>
    <action>search_add</action>
  </property>
  <property id="P4F" title="has_time-span">
    <table id="date"/>
    <order>1</order>
    <hides>
      <field>end_year</field>
      <field>end_month</field>
      <field>end_day</field>
      <field>end_comment</field>
      <field>end_extension</field>
    </hides>
    <action>input</action>
  </property>
</class>
```

Together these definitions form a template for the simplest possible person record. Figure 17 shows user-interface generated from the template. Saving the person record creates two equal records, the person and the birth, which are linked with a property. Both of them can be documented for exam- ple with texts and photographs. After the record is saved, it can be opened for editing, which allows attaching documents and add- ing new events and parts. When editing the record,

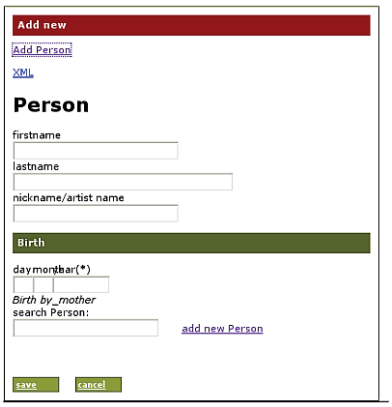


FIGURE 17 The graphical user-interface generated from the person template

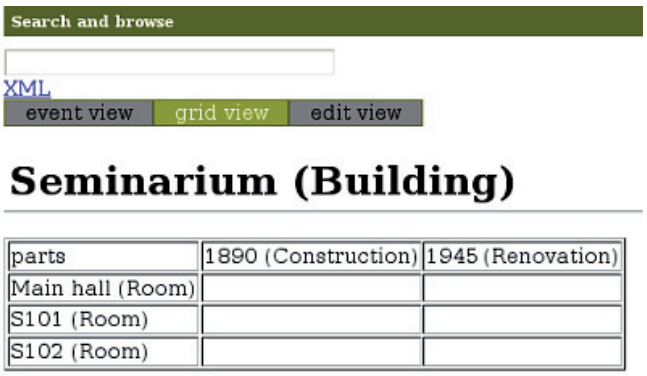


FIGURE 18 Events and parts of the building presented as a grid. Documents are not displayed.

possible events for that record are displayed in the menu from where they can be added.

Like stated earlier, using explicit events increases the resolution of documentation by splitting the history of a target to events and making those events themselves documentable. However, there must be also way to split target themselves to smaller units. In other words, express relations between targets and their parts. Therefore it is possible to define possible parts in a documentation template. Parts are individual records like any other record, but when for example a room is created as part of a building, it gets linked to the building with a property that expresses a part-of relationship between these two like shown in Figure 19.

The problematic aspect of this is how events behave with part-of relationships. If there was a renovation in the building, was there a renovation also in all parts of the building? From the documentation point of view a very practical solution was made. The event of the whole defines a time-period and by inheriting that event to all parts, it is possible to document what happened to parts dur-

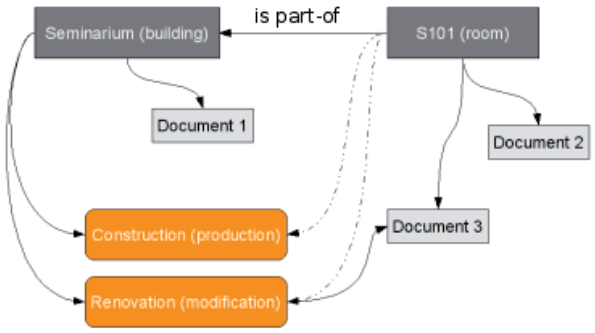


FIGURE 19 Inheritance of events

ing that time period. Figure 19 shows how events behave in this solution. The construction and the renovation events are inherited to the room (S101) from the building (Seminarium) but only the renovation event has a document (Document 3) that is related also to the room. This solution makes documentation faster because there is no need to repeat events for every part and it is easy to provide a table presentation of events and parts like shown in Figure 18.

Implementation

An open-source application called IDA-framework was developed in order to test ideas in practise. IDA-framework aims to be a simple and flexible tool for making semantically-aware, event-centric documentation. The core ontology of framework is based on CIDOC-CRM. However, the purpose is not to produce full CIDOC-CRM mappings but to provide a semantically meaningful base that follows CIDOC-CRM's principles and conventions. The application is divided in two parts: the server and the client. The server-side is written in PHP and it stores the data and takes care of application logic. The server uses relational database through MDB2-abstraction layer. The client is written with Javascript (AJAX) and it is responsible of creating graphical user-interface. Communication between the server and the client is done with XML- files. Using Ajax in the client-side makes overall user interface very flexible. User interface can be built in any layout with just a few lines of Javascript. The same data can be displayed differently just by modifying the script that generates the visual display in browser. It also makes possible to embed database to any website like for example Google Maps.

acknowledgements

This work has been funded by Department of Art and Culture studies in University of Jyväskylä and Finnish Cultural Foundation.

PIII

TOWARDS SEMANTIC MODELLING OF CULTURAL HISTORICAL DATA

by

Ari Häyrinen 2010

Information Modelling and Knowledge Bases XXII

Towards Semantic Modelling of Cultural Historical Data

In this paper a practical method is presented for creating documentation of cultural historical targets using an event-centric core ontology. By using semantic documentation templates and an XML-based query language, a domain specific documentation model can be created and flexible user interfaces can be built easily for accessing and editing the documentation. Keywords. ontologies, cultural historical documentation, information retrieval

Introduction

The most challenging feature of cultural historical data is its great variety. This variety sets challenges for documentation, information retrieval and user interface design in cultural historical data systems. Also, lately the question of information integration and especially that of cross-cultural information exchange have been raised in organisations such as museums, libraries and archives that collect and maintain cultural historical data. To solve some important problems in the field of cultural historical documentation(Bekiari et al., 2005), semantic technologies have been introduced. Conceptual Reference Model (later CRM) is a formal ontology intended to facilitate integration, mediation and interchange of heterogeneous cultural heritage information(Crofts et al., 2006). The model was created by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) on empirical bases from real-world datasets which reflect the special needs of the cultural historical field. CRM also has the status of being an ISO-standard(ISO, 2006). In this paper, an event-centric, CRM-based method for modelling cultural historical data is presented. The method is designed to help cultural historical documentation work by providing tools for semantically aware documentation of cultural historical items and events. In order to demonstrate method, a software called IDA-framework was implemented.

1. CIDOC-CRM and event-centric documentation

CIDOC-CRM consists of 86 classes and 143 properties, and it is meant to be extended by users for more specific domains. An important aspect of the CRM class hierarchy is separation between temporal entities and persistent items (Figure 20). It allows modelling of history as events, with actors participating in those events. Events can, for example, produce or modify persistent items, or persistent items can be used in the events.

1.1.Event-centric model of history

The tradition of documentation in the cultural historical field has been very item-centric(Cameron, 2007). The documentation is organised around physical items, which makes documentation of immaterial objects challenging or even impossible. In the event- centric approach, the root of documentation process is not nec-

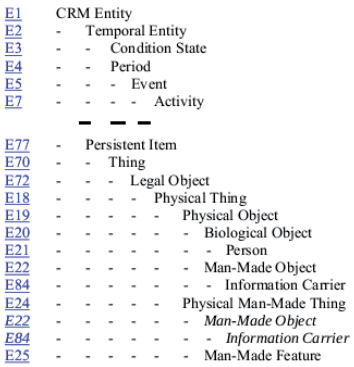


FIGURE 20 A partial CICOC-CRM class hierarchy

essarily a physical item. For example, when a hand-painted painting is modelled according to CIDOC-CRM, the painting has no direct property called author or artists. Instead, the painting is said to have been produced by a production event, which is carried out by one or more persons. Similarly, if painting is sold, damaged, or restored, these processes are modelled as individual events that affect the state of the painting.

The event-centric approach has several advantages compared to traditional, item- based approaches. First, events provide a semantically meaningful way of describing links between physical things and actions of human beings(Doerr and Kritsotaki, 2006). Second, the event- centric model allows a very flexible structure for an individual record. Events describe the history of an item, and new events can be added at any time. In practise, this means that the related documentation can be very detailed or just on a general level and that the level of details can be decided by the person creating the documentation. Therefore, documentation can be constructed based on the qualities of the target of documentation instead of some rigid structure.

Using explicit events also simplifies data structure design, because events have a common structure: someone did something somewhere during a certain time period. With this formulation, it is possible to represent, for example, a creation of an artwork, a construction of a building or having a scientific seminar. Because events are individual records, they split documentation into smaller units, therefore making it semantically more precise and more accessible.

The third, and, from the perspective of documentation, very important benefit is that using explicit events in documentation makes events themselves documentable as units. For example, a design process of a building or restoration of a painting can be documented as an individual record. In the traditional item-centric approach this would need specific fields for every event type. To be able to define the cultural context of an item, it is beneficial if the cultural object can be separated from the physical carrier object(s). For example, in the case of architectural drawings, it can be said that the immaterial architectural design is carried out by physical drawings. This way the actual design can be documented inde-

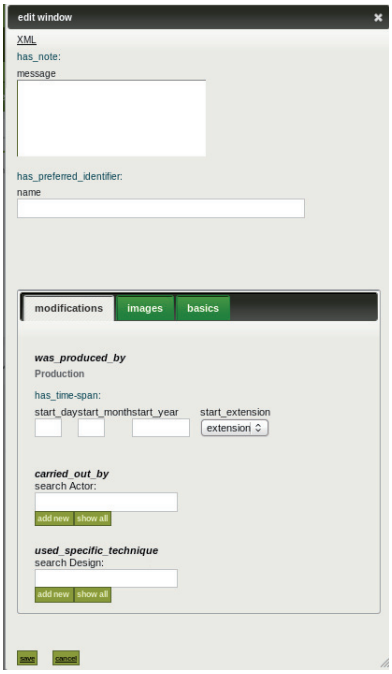


FIGURE 21 User interface created based on the documentation template

pendently from the physical documents. This separation also helps when modelling the relation between industrially made objects and conceptual designs. Industrially made furniture is produced by following a certain procedure or design. The design is documented only once even if there are several physical copies of the furniture.

However, sometimes it is easier to use a shorter path and model the relations in a shorter way. CIDOC-CRM also allows more simple modelling. For example, one way to model the relation between an architectural drawing and its target is to say that the drawing shows a building. Although this is not semantically accurate, since the drawing illustrates the design of the building and not the building itself, it can be useful when the initial data does not contain the necessary information for the design or when there are no resources to make full mappings.

2. IDA-framework

IDA-framework is a simple and flexible tool for making semantically-aware, event-centric documentation(Häyrinen, 2009). The core ontology of the framework is based on CIDOC- CRM. However, the purpose is not to produce full CIDOC-CRM mappings but to provide a semantically meaningful base that follows CIDOC-CRM's principles and conventions. IDA-framework is aimed specially for small memory organisations like local museums and highly specialized museums whose collections are densely linked. IDA-framework uses events explicitly in its doc-

umentation model. The complexity of the CIDOC-CRM is hidden from the end user, and the domain specific documentation structure is defined by semantic documentation templates(Häyrinen, 2008).

2.1.Semantic documentation template

Since CRM does not suggest what to document in any specific case, there must be a mechanism that guides users in making new records(Doerr and Iorizzo, 2008). The idea of a semantic documentation template is to provide a record-type specific documentation frame that can be defined by the organisation responsible for documentation. The template defines a typical case for a record, including default properties and default events. More precisely, the documentation template maps parts of thesauri to a partial domain ontology that is build on top of CIDOC-CRM.

```
<Building >
  <is_identified_by table="appellation" required="1" >
    <name required="0" />
  </is_identified_by >
  <has_note table="note" required="0" action="input" >
    <message required="0" width="200" />
  </has_note >
  <was_produced_by class="Production" required="1" >
    <Production >
      <has_time-span table="time_span" required="1" action="input" >
        <start_day required="0" />
        <start_month required="0" />
        <start_year required="1" width="5" />
      </has_time-span >
      <carried_out_by class="Actor" required="0" action="search_create" />
      <used_specific_technique class="Architectural_Design" required="0" />
    </Production >
  </was_produced_by >
  <has_current_location class="Place" />
</Building >
```

The documentation template does not define any user interface elements. A user interface can be freely created for the template concerned. Figure 21 shows an input form generated by Javascript in the browser. The template helps the user to create an initial target that can be later further refined.

2.2.Semantic query language

For performance reasons, IDA-framework is implemented with the help of a relational database system which excludes the use of Xquery or SPARQL as a query language.

The native query language for relational databases is SQL. While SQL is a well-known and established query language, it is also complex and it requires information about the internal data structure. Another problem with SQL is that queries can get very complex with the recursive database structures that IDA-framework uses. For this project, a new query language called IDA-QL was developed. IDA-QL is very verbal, and it should be meaningful without any prior knowledge about query languages. The following query will give the IDs of all

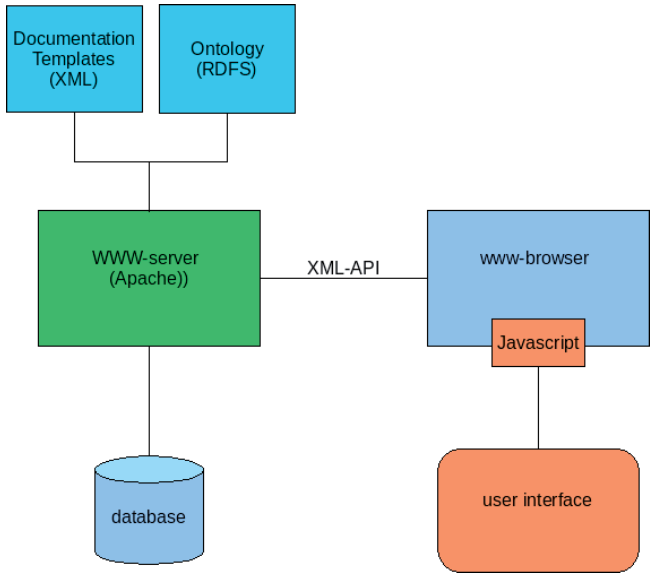


FIGURE 22 IDA-framework system architecture

persons who were born in 1953 and have "Lin*" in their name ("Linus" or "Lintunen", for example):

```
<?xml version="1.0" encoding="UTF-8" ?>
<search>
  <Person>
    <is_identified_by>
      <name>Lin*</name>
    </is_identified_by>
    <was_born>
      <Birth>
        <has_time-span>
          <start_year>1953</start_year>
        </has_time-span>
      </Birth>
    </was_born>
  </Person>
</search>
```

The purpose of IDA-QL is to provide platform independent information retrieval and data manipulation language for cultural historical data. It is a meta query language that is translated to SQL queries. IDA-QL is based on XML, and it operates on classes and properties defined by the ontology. This also means that if the ontology is translated to another language, then also the query language gets translated.

2.3.Implementation

IDA-framework server is written in PHP and it uses a relational database through the MDB2-abstraction layer. The ontology is defined by an RDFS file, and the documentation templates are described in an XML file. Communication between the server and the client is done with XML API. Client applications can be written

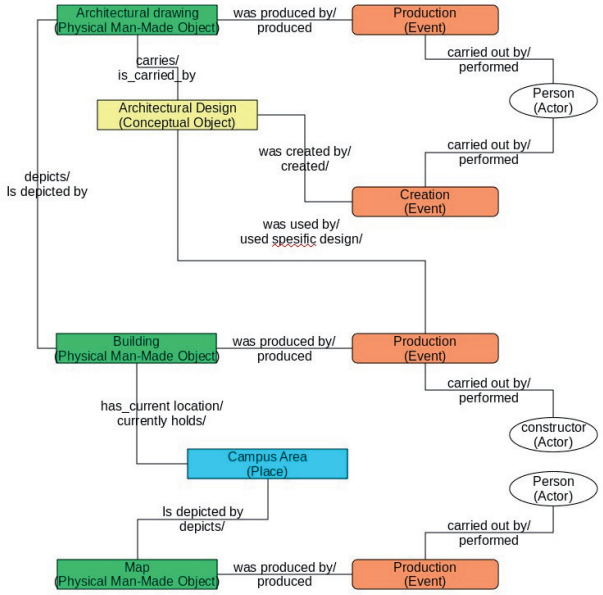


FIGURE 23 Maps and architectural drawings mapped to CRM.

with any programming language as long it can send http calls and is able to parse XML files. The overall implementation is very simple. There are fewer than 6000 lines of PHP code in the server application.

3. Case study: University Museum of Jyväskylä

Demonstrations were made with the collections of the University Museum of Jyväskylä. Two kinds of items were selected from the collections: maps and architectural drawings. There were about 200 maps and 100 architectural drawings in the museum's collection at the time of the experiment. The museum also has a collection management tool called DUO. Direct export form DUO to IDA-framework was not possible because of the differences in the documentation models. Instead, data was exported from DUO to an HTML file, which was then parsed with Javascript. The input was then manually validated and added to IDA-framework.

3.1.Domain Analysis based on CIDOC-CRM

A map depicts some geographical region produced by an individual person or organisation. In terms of CIDOC-CRM, map is a Physical Man-made Object produced by Production event and the Production event is carried out by an Actor. In this case, a simple model was used for modelling the authorship of the map. The producer of the physical map is not necessarily the creator of the map, but in

this case it is the assumption in the initial data. In addition, there is a possibility to define several map subtypes, a road map or a tourist map, among them. As stated earlier, a simple way to model a relation between an architectural drawing and its target is to say that the drawing shows a certain building. While useful in some cases, this model is not semantically accurate. The architectural drawing does not actually present the building, it merely presents the architectural design, which is an immaterial object. This relationship can be modeled with the conceptual class Architectural Design that is carried by the physical object – the drawing. In this case, a class called Architectural Design was derived from the CRM class Design_or_Procedure. This design was then linked to a production event of the building (Figure 23). This kind of modelling solves also the problem of how to document designs that were never carried out, in other words, plans that do not represent any physical building. The instance of Architectural Design exists even if there are no buildings constructed by following the design. The concept of architectural design also serves to organise drawings, because individual documents can be grouped as carriers of a single, named design.

3.2. Demo Applications

In the map application, the user can browse through maps by region, by map maker or by map type. The contents of these navigation panels are created by IDA-QL-queries. The following query gives the names of all persons and organisations who have produced a map:

```
<search>
  <Actor>
    <performed>
      <Production>
        <has_produced>
          <Map/>
        </has_produced>
      </Production>
    </performed>
  </Actor>
</search>
```

The query returns a list of actors. This list is then parsed by Javascript and displayed in a webpage. When the user clicks a certain name in the list, the following query is used to retrieve information:

```
<search result='result'>
  <Map>
    <was_produced_by>
      <Production>
        <carried_out_by>
          <Actor id='1' />
        </carried_out_by>
      </Production>
    </was_produced_by>
  </Map>
  <result>
    <depicts/>
    <has_type/>
    <was_produced_by>
      <carried_out_by />
    </was_produced_by>
  </result>
```

```
</search>
```

The query selects all the maps produced by the actor in question. The properties inside the result tag define what properties are included in the response XML. The XML response is then parsed with Javascript and the actual display is rendered. Architectural drawings can be viewed by buildings, by designs, by architects and by campus areas. Since this application is about architectural drawings, not about buildings, only the buildings that are presented in architectural drawings should be listed. The following query returns the list of buildings that are depicted by an architectural plan:

```
<Building>
  <is_depicted_by>
    <Architectural_Plan />
  </is_depicted_by>
</Building>
```

There is no concept of architect in the ontology used. However, the concept of architect can be defined by an IDA-QL query as a person who has produced a architectural drawing and thus a list of architects can be created.

4. Discussion

The case material was quite limited, and therefore it is too early to draw final conclusions. However, the material shows that the method can be used successfully in the field of cultural historical data, and a semantically rich Web 2.0 application can be built entirely with Javascript using an XML interface provided by IDA-framework. Modifications of the data structure do not require changes in the application code, which makes the system more flexible than other relational databases. The main focus of the software development was in the server side and on developing a flexible way to build user interfaces. Therefore, no actual usability tests were conducted at this point.

5. Conclusion

A practical method for CRM based documentation of cultural historical targets was presented. By hiding the complexity of the ontology with documentation templates, the system adapts semantic technologies in the field of cultural historical documentation. With a simple query language, it is possible to build flexible and browseable user interfaces without expertise needed for other semantic query languages.

Acknowledgements

This work has been funded by the Department of Art and Culture studies at the University of Jyväskylä, by the Finnish Cultural Foundation, and by Kone Foundation.

PIV

GOOGLE, WORLD'S LARGEST MUSEUM?

by

Ari Häyrinen 2011

CIDOC 2011, Sibiu, Romania

Google, world's largest museum?

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.

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 (Online Computer Library Center, 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, 2008a)

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 practise 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

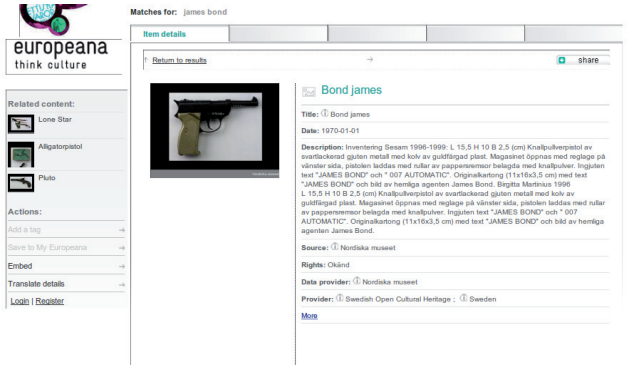


FIGURE 24 The gun in the Nordiska Museet's online database.

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, 2011a). 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.

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

FIGURE 25 James Bond cap gun on Ebay

"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].

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, s. a.).

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 ICT 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 (National Board of Antiquities, s. a.). 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, s. a.) 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.

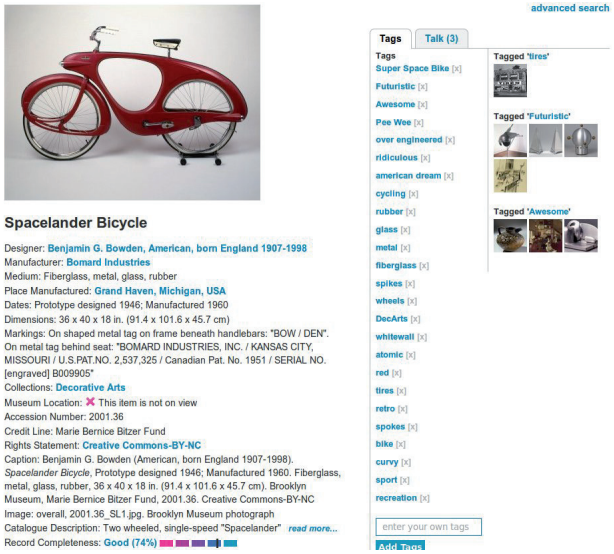


FIGURE 26 Bowden Spacelander in the Brooklyn Museum's online database.

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, s. 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, 1985)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 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, s. a.), 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.

Maretele Maimute 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 . 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 Maretele maimute 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.

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 "Maretele maimute" 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.

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

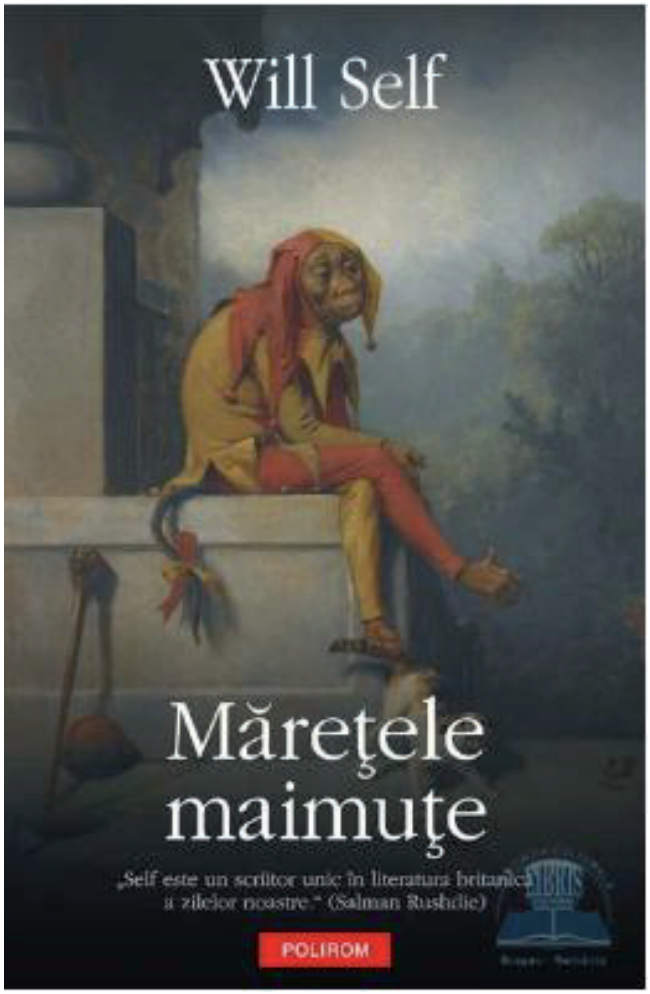


FIGURE 27 The cover of the Romanian edition of the Great Apes by Will Self

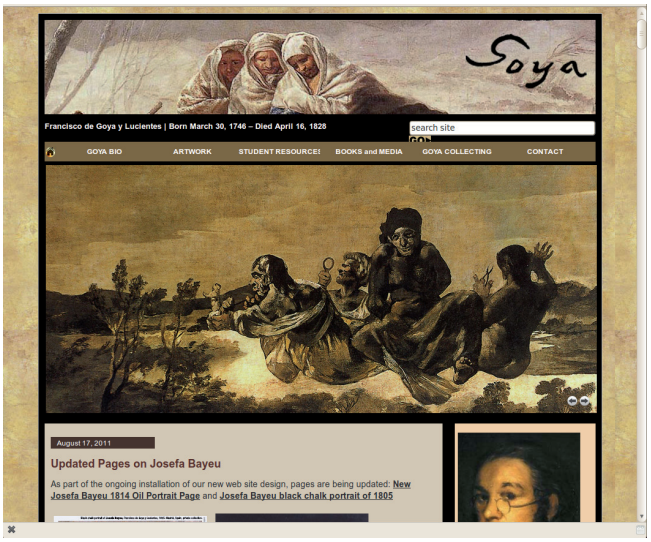


FIGURE 28 The Goya site at http://eeweems.com/goya/

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, s. a.). 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 archive in SIRIS . 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

FIGURE 29 An image in the Smithsonian Snapshot without the source information of the image.

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

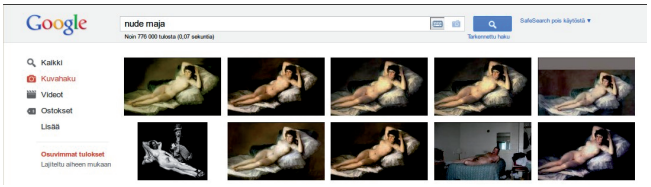


FIGURE 30 Goya's Nude Maja in the Google's image search with SafeSearch off. Note the nude male.

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 and 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. (Simons, 2008)

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 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,

FIGURE 31 Watermarked example images from kuvakokoelmat.fi (left) and from arjenhistoria.fi (right).

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 and 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 them 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.

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.

JYVÄSKYLÄ STUDIES IN HUMANITIES

1 KOSTIAINEN, EMMA, Viestintä ammattiosaamisen ulottuvuutena. - Communication as a dimension of vocational competence. 305 p. Summary 4 p. 2003.

2 SEPPÄLÄ, ANTTI, Todellisuutta kuvaamassa – todellisuutta tuottamassa. Työ ja koti television ja vähän radionkin uutisissa. - Describing reality – producing reality. Discourses of work and home in television and on a small scale in radio news. 211 p. Summary 3 p. 2003.

3 GERLANDER, MAIJA, Jännitteet lääkärin ja potilaan välisessä viestintäsuhteessa. - Tensions in the doctor-patient communication and relationship. 228 p. Summary 6 p. 2003.

4 LEHIKOINEN, TAISTO, Religious media theory - Understanding mediated faith and christian applications of modern media. - Uskonnollinen mediateoria: Modernin median kristilliset sovellukset. 341 p. Summary 5 p. 2003.

5 JARVA, VESA, Venäläisperäisyys ja ekspressiivisyys suomen murteiden sanastossa. - Russian influence and expressivity in the lexicon of Finnish dialects. 215 p. 6 p. 2003.

6 USKALI, TURO, "Älä kirjoita itseäsi ulos" Suomalaisen Moskovan-kirjeenvaihtajuuden alkutaival 1957–1975. - "Do not write yourself out" The beginning of the Finnish Moscow-correspondency in 1957–1975. 484 p. Summary 4 p. 2003.

7 VALKONEN, TARJA, Puheviestintätaitojen arviointi. Näkökulmia lukioikäisten esiintymis- ja ryhmätaitoihin. - Assessing speech communication skills. Perspectives on presentation and group communication skills among upper secondary school students. 310 p. Summary 7 p. 2003.

8 TAMPERE, KAJA, Public relations in a transition society 1989-2002. Using a stakeholder approach in organisational communications and relation analyses. 137 p. 2003.

9 EEROLA, TUOMAS, The dynamics of musical expectancy. Cross-cultural and statistical approaches to melodic expectations. - Musiikillisten odotusten tarkastelu kulttuurienvälisten vertailujen ja tilastollisten mallien avulla. 84 p. (277 p.) Yhteenveto 2 p. 2003.

10 PAANANEN, PIRKKO, Monta polkua musiikkiin. Tonaalisen musiikin perusrakenteiden kehittyminen musiikin tuottamis- ja improvisaatiotehtävissä ikävuosina 6–11. - Many paths to music. The development of basic structures of tonal music in music production and improvisation at the age of 6–11 years. 235 p. Summary 4 p. 2003.

11 LAAKSAMO, JOUKO, Musiikillisten karakterien metamorfoosi. Transformaatio- ja metamorfoosiprosessit Usko Meriläisen tuotannossa vuosina 1963–86. - "Metamorphosis of musical characters". Transformation and metamorphosis processes in the works of Usko Meriläinen during 1963–86. 307 p. Summary 3 p. 2004.

12 RAUTIO, RIITTA, *Fortspinnungstypus* Revisited. Schemata and prototypical features in J. S. Bach's Minor-Key Cantata Aria Introductions. - Uusi katsaus kehitysmuotoon. Skeemat ja prototyyppiset piirteet J. S. Bachin kantaattien molliaarioiden alkusoitoissa. 238 p. Yhteenveto 3 p. 2004.

13 MÄNTYLÄ, KATJA, Idioms and language users: the effect of the characteristics of idioms on their recognition and interpretation by native and non-native speakers of English. - Idiomien ominaisuuksien vaikutus englannin idiomien ymmärtämiseen ja tulkintaan syntyperäisten ja suomea äidinkielenään puhuvien näkökulmasta. 239 p. Yhteenveto 3 p. 2004.

14 MIKKONEN, YRJÖ, On conceptualization of music. Applying systemic approach to musicological concepts, with practical examples of music theory and analysis. - Musiikin käsitteellistämisestä. Systeemisen tarkastelutavan soveltaminen musikologisiin käsitteisiin sekä käytännön esimerkkejä musiikin teoriasta ja analyysistä. 294 p. Yhteenveto 10 p. 2004.

15 HOLM, JAN-MARKUS, Virtual violin in the digital domain. Physical modeling and model-based sound synthesis of violin and its interactive application in virtual environment. - Virtuaalinen viulu digitaalisella alueella. Viulun fysikaalinen mallintaminen ja mallipohjainen äänisynteesi sekä sen vuorovaikutteinen soveltaminen virtuaalitodellisuus ympäristössä. 74 p. (123 p.) Yhteenveto 1 p. 2004.

16 KEMP, CHRIS, Towards the holistic interpretation of musical genre classification. - Kohti musiikin genreluokituksen kokonaisvaltaista tulkintaa. 302 p. Yhteenveto 1 p. 2004.

17 LEINONEN, KARI, Finlandssvenskt sje-, tje- och s-ljud i kontrastiv belysning. 274 p. Yhteenveto 4 p. 2004.

18 MÄKINEN, EEVA, Pianisti cembalistina. Cembalotekniikka cembalonsoittoa aloittavan pianistin ongelmana. - The Pianist as cembalist. Adapting to harpsichord technique as a problem for pianists beginning on the harpsichord. 189 p. Summary 4 p. 2004.

19 KINNUNEN, MAURI, Herätysliike kahden kulttuurin rajalla. Lestadiolaisuus Karjalassa 1870-1939. - The Conviction on the boundary of two cultures. Laestadianism in Karelia in 1870-1939. 591 p. Summary 9 p. 2004.

20 Лилия Сиберг, "БЕЛЫЕ ЛИЛИИ". ГЕНЕЗИС ФИНСКОГО МИФА В БОЛГАРИИ. РОЛЬ РУССКОГО ФЕННОИЛЬСТВА. ФИНСКО-БОЛГАРСКИЕ КОНТАКТЫ И ПОСРЕДНИКИ С КОНЦА XIX ДО КОНЦА XX ВЕКА. 284 с. - "Belye lilii". Genezis finskogo mifa v Bolgarii. Rol' russkogo fennoil'stva. Finsko-bolgarskie kontakty i posredniki s konca XIX do konca XX veka. 284 p. Yhteenveto 2 p. 2004.

21 FUCHS, BERTOLD, Phonetische Aspekte einer Didaktik der Finnischen Gebärdensprache als Fremdsprache. - Suomalainen viittomakieli vieraana kielenä. Didaktinen fonetiikka. 476 p. Yhteenveto 14 p. 2004.

22 JÄÄSKELÄINEN, PETRI, Instrumentatiivisuus ja nykysuomen verbinjohto. Semanttinen tutkimus. - Instrumentality and verb derivation in Finnish. A semantic study. 504 p. Summary 5 p. 2004.

23 MERTANEN TOMI, Kahdentoista markan kapina? Vuoden 1956 yleislakko Suomessa. - A Rebellion for Twelve Marks? – The General Strike of 1956 in Finland. 399 p. Summary 10 p. 2004.

24 MALHERBE, JEAN-YVES, L'œuvre de fiction en prose de Marcel Thiry : une lecture d'inaboutissements. 353 p. Yhteenveto 1 p. 2004.

25 KUHNA, MATTI, Kahden maailman välissä. Marko Tapion *Arktinen hysteria* Väinö Linnan haastajana. - Between two worlds. Marko Tapio's Arktinen hysteria as a challenger to Väinö Linna. 307p. Summary 2 p. 2004.

26 VALTONEN, HELI, Minäkuvat, arvot ja mentaliteetit. Tutkimus 1900-luvun alussa syntyneiden toimihenkilönaisten omaelämäkerroista. - Self-images, values and mentalities. An autobiographical study of white collar women in twentieth century Finland. 272 p. Summary 6 p. 2004.

27 PUSZTAI, BERTALAN, Religious tourists. Constructing authentic experiences in late modern hungarian catholicism. - Uskontoturistit. Autenttisen elämyksen rakentaminen myöhäismodernissa unkarilaisessa katolisuudessa. 256 p. Yhteenveto 9 p. Summary in Hungarian 9 p. 2004.

28 PÄÄJOKI, TARJA, Taide kulttuurisena kohtaamispaikkana taidekavatuksessa. - The arts as a place of cultural encounters in arts education. 125 p. Summary 3 p. 2004.

29 JUPPI, PIRITA, "Keitä me olemme? Mitä me haluamme?" Eläinoikeusliike määrittelykamppailun, marginalisoinnin ja moraalisen paniikin kohteena suomalaisessa sanomalehdistössä. - "Who are we? What do we want?" The animal rights movement as an object of discursive struggle, marginalization and moral panic in Finnish newspapers. 315 p. Summary 6 p. 2004.

30 HOLMBERG, JUKKA, Etusivun politiikkaa. Yhteiskunnallisten toimijoiden representointi suomalaisissa sanomalehtiuutisissa 1987–2003. - Front page politics. Representation of societal actors in Finnish newspapers' news articles in 1987-2003. 291 p. Summary 2 p. 2004.

31 LAGERBLOM, KIMMO, Kaukana Kainuussa, valtaväylän varrella. Etnologinen tutkimus Kontiomäen rautatieläisyhteisön elinkaaresta 1950 – 1972. - Far, far away, nearby a main passage. An ethnological study of the life spans of Kontiomäki railtown 1950 – 1972. 407 p. Summary 2 p. 2004.

32 HAKAMÄKI, LEENA, Scaffolded assistance provided by an EFL teacher during whole-class interaction. - Vieraan kielen opettajan antama oikea-aikainen tuki luokkahuoneessa. 331 p. Yhteenveto 7 p. 2005.

33 VIERGUTZ, GUDRUN, Beiträge zur Geschichte des Musikunterrichts an den Gelehrtenschulen der östlichen Ostseeregion im 16. und 17. Jahrhundert. - Latinankoulujen musiikinopetuksen historiasta itäisen Itämeren rannikkokaupungeissa 1500- ja 1600-luvuilla. 211 p. Yhteenveto 9 p. 2005.

34 NIKULA, KAISU, Zur Umsetzung deutscher Lyrik in finnische Musik am Beispiel Rainer Maria Rilke und Einojuhani Rautavaara. - Saksalainen runous suomalaisessa musiikissa, esimerkkinä Rainer Maria Rilke ja Einojuhani Rautavaara. 304 p. Yhteenveto 6 p. 2005.

35 SYVÄNEN, KARI, Vastatunteiden dynamiikka musiikkiterapiassa. - Counter emotions dynamics in music therapy. 186 p. Summary 4 p. 2005.

36 ELORANTA, JARI & OJALA, JARI (eds), East-West trade and the cold war. 235 p. 2005.

37 HILTUNEN, KAISA, Images of time, thought and emotions: Narration and the spectator's experience in Krzysztof Kieslowski 's late fiction films. - Ajan, ajattelun ja tunteiden kuvia. Kerronta ja katsojan kokemus Krzysztof Kieslowskin myöhäisfiktiossa. 203 p. Yhteenveto 5 p. 2005.

38 AHONEN, KALEVI, From sugar triangle to cotton triangle. Trade and shipping between America and Baltic Russia, 1783-1860. 572 p. Yhteenveto 9 p. 2005.

39 UTRIAINEN, JAANA, A gestalt music analysis. Philosophical theory, method, and analysis of Iegor Reznikoff's compositions. - Hahmoperustainen musiikkianalyysi. Hahmofilosofinen teoria, metodi ja musiikkianalyysi Iégor Reznikoffin sävellyksistä. 222 p. Yhteenveto 3 p. 2005.

40 MURTORINNE, ANNAMARI, *Tuskan hauskaa*! Tavoitteena tiedostava kirjoittaminen. Kirjoittamisprosessi peruskoulun yhdeksännellä luokalla. - Painfully fun! Towards reflective writing process. 338 p. 2005.

41 TUNTURI, ANNA-RIITTA, Der Pikareske Roman als Katalysator in Geschichtlichen Abläufen. Erzählerische Kommunikationsmodelle in *Das Leben des Lazarillo von Tormes*, bei Thomas Mann und in EinigenFinnischen Romanen. 183 p. 2005.

42 LUOMA-AHO, VILMA, Faith-holders as Social Capital of Finnish Public Organisations. - Luottojoukot – Suomalaisten julkisten organisaatioiden sosiaalista pääomaa. 368 p. Yhteenveto 8 p. 2005.

63 RUUSKANEN, ESA, Viholliskuviin ja viranomaisiin vetoamalla vaiennetut työväentalot. Kuinka Pohjois-Savon Lapuan liike sai nimismiehet ja maaherran sulkemaan 59 kommunistista työväentaloa Pohjois-Savossa vuosina 1930–1932. - The workers' halls closed by scare-mongering and the use of special powers by the authorities. 248 p. Summary 5 p. 2006.

64 VARDJA, MERIKE, Tegelaskategooriad ja tegelase kujutamise vahendid Väinö Linna romaanis "Tundmatu sõdur". - Character categories and the means of character representation in Väinö Linna's Novel *The Unknown Soldier*. 208 p. Summary 3 p. 2006.

65 TAKÁTS, JÓZSEF, Módszertani berek. Írások az irodalomtörténet-írásról. - The Grove of Methodology. Writings on Literary Historiography. 164 p. Summary 3 p. 2006.

66 MIKKOLA, LEENA, Tuen merkitykset potilaan ja hoitajan vuorovaikutuksessa. - Meanings of social support in patient-nurse interaction. 260 p. Summary 3 p. 2006.

67 SAARIKALLIO, SUVI, Music as mood regulation in adolescence. - Musiikki nuorten tunteiden säätelynä. 46 p. (119 p.) Yhteenveto 2 p. 2007.

68 HUJANEN, ERKKI, Lukijakunnan rajamailla. Sanomalehden muuttuvat merkitykset arjessa. - On the fringes of readership. The changing meanings of newspaper in everyday life. 296 p. Summary 4 p. 2007.

69 TUOKKO, EEVA, Mille tasolle perusopetuksen englannin opiskelussa päästään? Perusopetuksen päättövaiheen kansallisen arvioinnin 1999 eurooppalaisen viitekehyksen taitotasoihin linkitetyt tulokset. - What level do pupils reach in English at the end of the comprehensive school? National assessment results linked to the common European framework. 338 p. Summary 7 p. Sammanfattning 1 p. Tiivistelmä 1 p. 2007.

70 TUIKKA, TIMO, "Kekkosen konstit". Urho Kekkosen historia- ja politiikkakäsitykset teoriasta käytäntöön 1933–1981. - "Kekkonen´s way". Urho Kekkonen's conceptions of history and politics from theory to practice, 1933–1981 413 p. Summary 3 p. 2007.

71 Humanistista kirjoa. 145 s. 2007.

72 NIEMINEN, LEA, A complex case: a morphosyntactic approach to complexity in early child language. 296 p. Tiivistelmä 7 p. 2007.

73 TORVELAINEN, PÄIVI, Kaksivuotiaiden lasten fonologisen kehityksen variaatio. Puheen ymmärrettävyyden sekä sananmuotojen tavoittelun ja tuottamisen tarkastelu. - Variation in phonological development of two-year-old Finnish children. A study of speech intelligibility and attempting and production of words. 220 p. Summary 10 p. 2007.

74 SIITONEN, MARKO, Social interaction in online multiplayer communities. - Vuorovaikutus verkkopeliyhteisöissä. 235 p. Yhteenveto 5 p. 2007.

75 STJERNVALL-JÄRVI, BIRGITTA, Kartanoarkkitehtuuri osana Tandefelt-suvun elämäntapaa. - Manor house architecture as part of the Tandefelt family´s lifestyle. 231 p. 2007.

76 SULKUNEN, SARI, Text authenticity in international reading literacy assessment. Focusing on PISA 2000. - Tekstien autenttisuus kansainvälisissä lukutaidon arviointitutkimuksissa: PISA 2000. 227 p. Tiivistelmä 6 p. 2007.

77 KŐSZEGHY, PÉTER, Magyar Alkibiadés. Balassi Bálint élete. - The Hungarian Alcibiades. The life of Bálint Balass. 270 p. Summary 6 p. 2007.

78 MIKKONEN, SIMO, State composers and the red courtiers - Music, ideology, and politics in the Soviet 1930s - Valtion säveltäjiä ja punaisia hoviherroja. Musiikki, ideologia ja politiikka 1930-luvun Neuvostoliitossa. 336 p. Yhteenveto 4 p. 2007.

79 SIVUNEN, ANU, Vuorovaikutus, viestintäteknologia ja identifioituminen hajautetuissa tiimeissä. - Social interaction, communication technology and identification in virtual teams. 251 p. Summary 6 p. 2007.

80 LAPPI, TIINA-RIITTA, Neuvottelu tilan tulkinnoista. Etnologinen tutkimus sosiaalisen ja materiaalisen ympäristön vuorovaikutuksesta jyväskyläläisissä kaupunkipuhunnoissa. - Negotiating urban spatiality. An ethnological study on the interplay of social and material environment in urban narrations on Jyväskylä. 231 p. Summary 4 p. 2007.

81 HUHTAMÄKI, ULLA, "Heittäydy vapauteen". Avantgarde ja Kauko Lehtisen taiteen murros 1961–1965. - "Fling yourself into freedom!" The Avant-Garde and the artistic transition of Kauko Lehtinen over the period 1961–1965. 287 p. Summary 4 p. 2007.

82 KELA, MARIA, *Jumalan kasvot* suomeksi. Metaforisaatio ja erään uskonnollisen ilmauksen synty. - God's face in Finnish. Metaphorisation and the emergence of a religious expression. 275 p. Summary 5 p. 2007.

83 SAARINEN, TAINA, Quality on the move. Discursive construction of higher education policy from the perspective of quality. - Laatu liikkeessä. Korkeakoulupolitiikan diskursiivinen rakentuminen laadun näkökulmasta. 90 p. (176 p.) Yhteenveto 4 p. 2007.

84 MÄKILÄ, KIMMO, Tuhoa, tehoa ja tuhlausta. Helsingin Sanomien ja New York Timesin ydinaseuutisoinnin tarkastelua diskurssianalyyttisesta näkökulmasta 1945–1998.

- "Powerful, Useful and Wasteful". Discourses of Nuclear Weapons in the New York Times and Helsingin Sanomat 1945–1998. 337 p. Summary 7 p. 2007.

85 KANTANEN, HELENA, Stakeholder dialogue and regional engagement in the context of higher education. - Yliopistojen sidosryhmävuoropuhelu ja alueellinen sitoutuminen. 209 p. Yhteenveto 8 p. 2007.

86 ALMONKARI, MERJA, Jännittäminen opiskelun puheviestintätilanteissa. - Social anxiety in study-related communication situations. 204 p. Summary 4 p. 2007.

87 VALENTINI, CHIARA, Promoting the European Union. Comparative analysis of EU communication strategies in Finland and in Italy. 159 p. (282 p.) 2008.

88 PULKKINEN, HANNU, Uutisten arkkitehtuuri - Sanomalehden ulkoasun rakenteiden järjestys ja jousto. - The Architecture of news. Order and flexibility of newspaper design structures. 280 p. Yhteenveto 5 p. 2008.

89 MERILÄINEN, MERJA, Monenlaiset oppijat englanninkielisessä kielikylpyopetuksessa - rakennusaineita opetusjärjestelyjen tueksi. - Diverse Children in English Immersion: Tools for Supporting Teaching Arrangements. 197 p. 2008.

90 VARES, MARI, The question of Western Hungary/Burgenland, 1918-1923. A territorial question in the context of national and international policy. - Länsi-Unkarin/Burgenlandin kysymys 1918–1923. Aluekysymys kansallisen ja kansainvälisen politiikan kontekstissa. 328 p. Yhteenveto 8 p. 2008.

91 ALA-RUONA, ESA, Alkuarviointi kliinisenä käytäntönä psyykkisesti oireilevien asiakkaiden musiikkiterapiassa – strategioita, menetelmiä ja apukeinoja. – Initial assessment as a clinical procedure in music therapy of clients with mental health problems – strategies, methods and tools. 155 p. 2008.

92 ORAVALA, JUHA, Kohti elokuvallista ajattelua. Virtuaalisen todellisen ontologia Gilles Deleuzen ja Jean-Luc Godardin elokuvakäsityksissä. - Towards cinematic thinking. The ontology of the virtually real in Gilles Deleuze's and Jean-Luc Godard's conceptions of cinema. 184 p. Summary 6 p. 2008.

93 KECSKEMÉTI, ISTVÁN, Papyruksesta megabitteihin. Arkisto- ja valokuvakokoelmien konservoinnin prosessin hallinta. - From papyrus to megabytes: Conservation management of archival and photographic collections. 277 p. 2008.

94 SUNI, MINNA, Toista kieltä vuorovaikutuksessa. Kielellisten resurssien jakaminen toisen kielen omaksumisen alkuvaiheessa. - Second language in interaction: sharing linguistic resources in the early stage of second language acquisition. 251 p. Summary 9 p. 2008.

95 N. PÁL, JÓZSEF, Modernség, progresszió, Ady Endre és az Ady–Rákosi vita. Egy konfliktusos eszmetörténeti pozíció természete és következményei. 203 p. Summary 3 p. 2008.

96 BARTIS, IMRE, „Az igazság ismérve az, hogy igaz". Etika és nemzeti identitás Sütő András Anyám könnyű álmot ígér című művében és annak recepciójában. 173 p. Summary 4 p. 2008.

97 RANTA-MEYER, TUIRE, Nulla dies sine linea. Avauksia Erkki Melartinin vaikutteisiin, verkostoihin ja vastaanottoon henkilö- ja reseptiohistoriallisena tutkimuksena. - *Nulla dies sine linea*: A biographical and reception-historical approach to Finnish composer Erkki Melartin. 68 p. Summary 6 p. 2008.

98 KOIVISTO, KEIJO, Itsenäisen Suomen kanta-aliupseeriston synty, koulutus, rekrytointitausta ja palvelusehdot. - The rise, education, the background of recruitment and conditions of service of the non-commissioned officers in independent Finland. 300 p. Summary 7 p. 2008.

99 KISS, MIKLÓS, Between narrative and cognitive approaches. Film theory of non-linearity applied to Hungarian movies. 198 p. 2008.

100 RUUSUNEN, AIMO, Todeksi uskottua. Kansandemokraattinen Neuvostoliitto-journalismi rajapinnan tulkkina vuosina1964–1973. - Believed to be true. Reporting on the USSR as interpretation of a boundary surface in pro-communist partisan journalism 1964–1973. 311 p. Summary 4 p. 2008.

101 HÄRMÄLÄ, MARITA, Riittääkö *Ett ögonblick* näytöksi merkonomilta edellytetystä kielitaidosta? Kielitaidon arviointi aikuisten näyttötutkinnoissa. – Is *Ett ögonblick* a sufficient demonstration of the language skills required in the qualification of business and administration? Language assessment in competence-based qualifications for adults. 318 p. Summary 4 p. 2008.

102 COELHO, JACQUES, The vision of the cyclops. From painting to video ways of seeing in the 20th century and through the eyes of Man Ray. 538 p. 2008.

103 BREWIS, KIELO, Stress in the multi-ethnic customer contacts of the Finnish civil servants: Developing critical pragmatic intercultural professionals. – Stressin kokemus suomalaisten viranomaisten monietnisissä asiakaskontakteissa: kriittis-pragmaattisen kulttuurienvälisen ammattitaidon kehittäminen. 299 p. Yhteenveto 4 p. 2008.

104 BELIK, ZHANNA, The Peshekhonovs' Workshop: The Heritage in Icon Painting. 239 p. [Russian]. Summary 7 p. 2008.

105 MOILANEN, LAURA-KRISTIINA, Talonpoikaisuus, säädyllisyys ja suomalaisuus 1800- ja 1900-lukujen vaihteen suomenkielisen proosan kertomana. – Peasant values, estate society and the Finnish in late nineteenth- and early

and early twentieth-century narrative literature. 208 p. Summary 3 p. 2008.

106 PÄÄRNILÄ, OSSI, Hengen hehkusta tietostrategioihin. Jyväskylän yliopiston humanistisen tiedekunnan viisi vuosikymmentä. 110 p. 2008.

107 KANGASNIEMI, JUKKA, Yksinäisyyden kokemisen avainkomponentit Yleisradion tekstitelevision Nuorten palstan kirjoituksissa. - The key components of the experience of loneliness on the Finnish Broadcasting Company's (YLE) teletext forum for adolescents. 388 p. 2008.

108 GAJDÓ, TAMÁS, Színháztörténeti metszetek a 19. század végétől a 20. század közepéig. - Segments of theatre history from the end of the 19th century to the middle of the 20th century. 246 p. Summary 2 p. 2008.

109 CATANI, JOHANNA, Yritystapahtuma kontekstina ja kulttuurisena kokemuksena. - Corporate event as context and cultural experience. 140 p. Summary 3 p. 2008.

110 MAHLAMÄKI-KAISTINEN, RIIKKA, Mätänevän velhon taidejulistus. Intertekstuaalisen ja -figuraalisen aineiston asema Apollinairen L'Enchanteur pourrissant teoksen tematiikassa ja symboliikassa. - Pamphlet of the rotten sorcerer. The themes and symbols that intertextuality and interfigurality raise in Apollinaire's prose work L'Enchanteur pourrissant. 235 p. Résumé 4 p. 2008.

111 PIETILÄ, JYRKI, Kirjoitus, juttu, tekstielementti. Suomalainen sanomalehtijournalismi juttutyyppien kehityksen valossa printtimedian vuosina 1771-2000. - Written Item, Story, Text Element. Finnish print journalism in the light of the development of journalistic genres during the period 1771-2000. 779 p. Summary 2 p. 2008.

112 SAUKKO, PÄIVI, Musiikkiterapian tavoitteet lapsen kuntoutusprosessissa. - The goals of music therapy in the child's rehabilitation process. 215 p. Summary 2 p. 2008.

113 LASSILA-MERISALO, MARIA, Faktan ja fiktion rajamailla. Kaunokirjallisen journalismin poetiikka suomalaisissa aikakauslehdissä. - On the borderline of fact and fiction. The poetics of literary journalism in Finnish magazines. 238 p. Summary 3 p. 2009.

114 KNUUTINEN, ULLA, Kulttuurihistoriallisten materiaalien menneisyys ja tulevaisuus. Konservoinnin materiaalitutkimuksen heritologiset funktiot. - The heritological functions of materials research of conservation. 157 p. (208 p.) 2009.

115 NIIRANEN, SUSANNA, «Miroir de mérite». Valeurs sociales, rôles et image de la femme dans les textes médiévaux des *trobairitz*. - "Arvokkuuden peili". Sosiaaliset arvot, roolit ja naiskuva keskiaikaisissa *trobairitz*-teksteissä. 267 p. Yhteenveto 4 p. 2009.

116 ARO, MARI, Speakers and doers. Polyphony and agency in children's beliefs about language learning. - Puhujat ja tekijät. Polyfonia ja agentiivisuus lasten kielenoppimiskäsityksissä. 184 p. Yhteenveto 5 p. 2009.

117 JANTUNEN, TOMMI, Tavu ja lause. Tutkimuksia kahden sekventiaalisen perusyksikön olemuksesta suomalaisessa viittomakielessä. - Syllable and sentence. Studies on the nature of two sequential basic units in Finnish Sign Language. 64 p. 2009.

118 SÄRKKÄ, TIMO, Hobson's Imperialism. A Study in Late-Victorian political thought. - J. A. Hobsonin imperialismi. 211 p. Yhteenveto 11 p. 2009.

119 LAIHONEN, PETTERI, Language ideologies in the Romanian Banat. Analysis of interviews and academic writings among the Hungarians and Germans. 51 p. (180 p) Yhteenveto 3 p. 2009.

120 MÁTYÁS, EMESE, Sprachlernspiele im DaF-Unterricht. Einblick in die Spielpraxis des finnischen und ungarischen Deutsch-als-Fremdsprache-Unterrichts in der gymnasialen Oberstufe sowie in die subjektiven Theorien der Lehrenden über den Einsatz von Sprachlernspielen. 399 p. 2009.

121 PARACZKY, ÁGNES, Näkeekö taitava muusikko sen minkä kuulee? Melodiadiktaatin ongelmat suomalaisessa ja unkarilaisessa taidemusiikin ammattikoulutuksessa. - Do accomplished musicians see what they hear? 164 p. Magyar nyelvü összefoglaló 15 p. Summary 4 p. 2009.

122 ELOMAA, EEVA, Oppikirja eläköön! Teoreettisia ja käytännön näkökohtia kielten oppimateriaalien uudistamiseen. - Cheers to the textbook! Theoretical and practical considerations on enchancing foreign language textbook design. 307 p. Zusammanfassung 1 p. 2009.

123 HELLE, ANNA, Jäljet sanoissa. Jälkistrukturalistisen kirjallisuuskäsityksen tulo 1980-luvun Suomeen. - Traces in the words. The advent of the poststructuralist conception of literature to Finland in the 1980s. 272 p. Summary 2 p. 2009.

124 PIMIÄ, TENHO ILARI, Tähtäin idässä. Suomalainen sukukansojen tutkimus toisessa maailmansodassa. - Setting sights on East Karelia: Finnish ethnology during the Second World War. 275 p. Summary 2 p. 2009.

125 VUORIO, KAIJA, Sanoma, lähettäjä, kulttuuri. Lehdistöhistorian tutkimustraditiot Suomessa ja median rakennemuutos. - Message, sender, culture. Traditions of research into the history of the press in Finland and structural change in the media. 107 p. 2009.

126 BENE, ADRIÁN Egyén és közösség. Jean-Paul Sartre *Critique de la raison dialectique* című műve a magyar recepció tükrében. - Individual and community. Jean-Paul Sartre's

165 Rautavuoma, Veera, Liberation exhibitions as a commemorative membrane of socialist Hungary. 251 p. Yhteenveto 3 p. 2011.
166 Lehtonen, Kimmo E., Rhetoric of the visual – metaphor in a still image. – Visuaalisen retoriikka – metafora still-kuvan tarkastelussa. 174 p. Yhteenveto 1 p. 2011.
167 Sarkamo, Ville, Karoliinien soturiarvot. Kunnian hallitsema maailmankuva Ruotsin valtakunnassa 1700-luvun alussa. – Carolean warrior values: an honour-dominated worldview in early-eighteenth-century Sweden. 216 p. Summary 11 p. 2011.
168 Rynkänen, Tatjana, Русскоязычные молодые иммигранты в Финляндии – интеграция в контексте обучения и овладения языком. – Russian-speaking immigrant adolescents in Finnish society – integration from the perspective of language and education. 258 p. Tiivistelmä 9 p. Summary 9 p. 2011.
169 Tiainen, Veikko, Vähentäjää vähentämässä. Tehdaspuu Oy puunhankkijana Suomessa. – Tehdaspuu Oy in Finnish Wood Procurement. 236 p. Summary 5 p. 2011.
170 Stolp, Marleena, Taidetta, vastustusta, leikkiä ja työtä? Lasten toimijuus 6–vuotiaiden teatteriprojektissa. – Art, resistance, play and work? Children's agency in a six-year-olds' theatre project. 79 p. (142 p.) 2011.
171 Cools, Carine, Relational dialectics in intercultural couples' relationships. – Kulttuurienvälisten parisuhteiden relationaalinen dialektiikka. 282 p. 2011.
172 Saario, Johanna, Yhteiskuntaopin kieliympäristö ja käsitteet – toisella kielellä opiskelevan haasteet ja tuen tarpeet. – The language environment and concepts in social studies – challenges and need of support for a second language learner. 290 p. Summary 7 p. 2012.
173 Alluri, Vinoo, Acoustic, neural, and perceptual correlates of polyphonic timbre. – Polyfonisen sointivärin hahmottamisen akustiset ja hermostolliset vastineet. 76 p. (141 p.) Yhteenveto 1 p. 2012.
174 Vuoskoski, Jonna Katariina, Emotions represented and induced by music: The role of individual differences. – Yksilöllisten erojen merkitys musiikillisten emootioiden havaitsemisessa ja kokemisessa. 58 p. (132 p.) Tiivistelmä 1 p. 2012.
175 Leinonen, Jukka, The beginning of the cold war as a phenomenon of realpolitik – U.S. secretary of state James F. Byrnes in the field of power politics 1945–1947. – Kylmän sodan synty reaalipoliittisena ilmiönä – James F. Byrnes suurvaltapolitiikan pelikentällä Jaltasta Stuttgartiin 1945–1947. 393 p. Yhteenveto 8 p. 2012.
176 Thompson, Marc, The application of motion capture to embodied music cognition research. - Liikkeenkaappausteknologian soveltaminen kehollisen musiikkikognition tutkimuksessa. 86 p. (165 p.) Yhteenveto 1 p. 2012.
177 Ferrer, rafael, The socially distributed cognition of musical timbre: a convergence of semantic, perceptual, and acoustic aspects. - Musiikillisen sointivärin jakautunut kognitio. 42 p. (156 p.) Yhteenveto 1 p. 2012.
178 Purhonen, Pipsa, Interpersonal communication competence and collaborative interaction in SME internationalization. 72 p. (196 p.) Yhteenveto 7 p. 2012.
179 Aira, Annaleena, Toimiva yhteistyö - työelämän vuorovaikutussuhteet, tiimit ja verkostot. - Successful collaboration - interpersonal relationships, teams and networks in working life. 182 p. Summary 4 p. 2012.
180 Huhtinen-Hildén, Laura, Kohti sensitiivistä musiikin opettamista - ammattitaidon ja opettajuuden rakentumisen polkuja. - Towards sensitive music teaching. Pathways to becoming a professional music educator. 300 p. Summary 2 p. 2012.
181 Jäntti, Saara, Bringing Madness Home. The Multiple Meanings of Home in Janet Frame's Faces in the Water, Bessie Head's A Question of Power and Lauren Slater's Prozac Diary. - Kodin monet merkitykset naisten hulluuskertomuksissa: Janet Framen Faces in the Water, Bessie Headin A Question of Power ja Lauren Slaterin Prozac Diary. 358 p. Yhteenveto 7 p. 2012.
182 Copp Jinkerson, Alicia, Socialization, language choice and belonging: Language norms in a first and second grade English medium class. - Sosiaalistuminen, kielen valinta ja jäsenyys: kielelliset normit englanninkielisessä alakoulun ensimmäisen ja toisen luokan opetuksessa Suomessa. 108 p. (197 p.) Yhteenveto 5 p. 2012.
183 Räihä, Antti, Jatkuvuus ja muutosten hallinta. Hamina ja Lappeenranta Ruotsin ja Venäjän alaisuudessa 1720−1760-luvuilla. - Continuity and the management of change: Hamina and Lappeenranta under Swedish and Russian rule from the 1720s to the 1760s. 279 p. Summary 12 p. 2012.
184 Miettinen, Helena, Phonological working memory and L2 knowledge: Finnish children learning English. - Fonologinen työmuisti ja vieraan kielen taito: Suomalaiset lapset englannin oppijoina. 187 p. Yhteenveto 4 p. 2012.
185 Fleisz, Katalin, Medialitás Krúdy Gyula prózájában. - Mediation In Krúdy Gyula's Prose. 204 p. Summary 2 p. 2012.
186 Karjalainen, Anu, Liikkuva ja muuttuva suomi. Diskursiivis-etnografinen tutkimus amerikansuomalaisten kielielämäkerroista. - The Finnish language in motion. An ethnographic and discourse analytic study of Finnish-American language biographies. 288 p. Summary 5 p. 2012.
187 Häyrinen, Ari, Open sourcing digital heritage. Digital surrogates, museums and knowledge. 134 p. (173 p.) Yhteenveto 3 p. 2012.