

Challenges of Diversity, Consistency, and Globality in Indexing of Local Archeological Artifacts[†]

Teija Oikarinen* and Terttu Kortelainen**

*Archaeology, Faculty of Humanities, P.O. Box 1000,
90014 University of Oulu, Finland, <teija.oikarinen@oulu.fi>

**Information Studies, Faculty of Humanities, P.O. Box 1000,
90014 University of Oulu, Finland, <terttu.kortelainen@oulu.fi>

Teija Oikarinen, MA archeology, holds a doctoral student position in the doctoral programme of Memornet, coordinated by the University of Tampere and financed by the Finnish Academy. The Memornet trains researchers from various fields of study to be capable of identifying changes in memory functions and to react to these. Her ongoing dissertation work focuses on digitalization of archeological data as local and global process, specifically on archeological excavation and the relation between archeological practices and data at the turning point of the digitalized eScience era. She has worked as a field manager in archeological excavations, as well.

Terttu Kortelainen, Ph.D., is University Lecturer of Information Studies at the University of Oulu, Finland. Her research interests are in informetric research, social media, and the evaluation of libraries. She has also supervised research projects focusing on usability of web services and evaluation of libraries. Her publications consist of four study books, a doctoral thesis, articles on informetrics, and articles on the study projects of the department. She is a member of the Advisory Committee of the Finnish Social Science Data Archive and the publication board of the Finnish Information Studies publication series.



Oikarinen, Teija, and Kortelainen, Terttu. **Challenges of Diversity, Consistency, and Globality in Indexing of Local Archeological Artifacts.** *Knowledge Organization.* 40(2), 123-135. 55 references.

ABSTRACT: We consider documents produced in archeological post-excavation analysis and re-raise a question of archeological cataloguing, which is a specific case in the context of global progress of digitalization in archeology. The catalogue of archeological artifacts from the excavation of the city of Jakobstad, Finland was analyzed through a content analysis. Quantitative analysis was conducted using SPSS statistical package, and the results are presented in figures and tables. The analysis was based on a qualitative definition of variables describing the archeological artifacts. The analysis shows that the catalogue of artifacts is mainly systematic, but the results also reveal non-uniformity in cataloguing. In the free description column, several categorizations were found that could be used in developing the structure of an archeological catalogue. Traditional cataloguing methods are still practiced in archeology, but these do not fulfill requirements of the future use of data. In this case, a vocabulary and a tool for cataloguing archeological artifacts would contribute to the development of cataloguing and future access of data. These devices should be flexible and support uniqueness of the artifacts. There exist tools and vocabularies for archeological cataloguing and these could be localized to fulfill the needs for the future digitalization of archeological data.

Received 11 December 2012; Revised 4 January 2013; Accepted 8 January 2013

[†] The authors thank the Finnish Cultural Foundation and the Memornet Doctoral Programme for supporting this work. The authors thank the editor and two anonymous referees for their insightful comments and suggestions as well.

1.0 Introduction

In archeology, field studies such as excavations are the most visible part of research. However, it is a short period compared to the post-excavation stage in which arti-

facts are classified and described on the basis of their archeological identification and interpretation such as use or function of objects and their dating. The catalogues and reports produced in the post-excavation stage are a very relevant part of an archeological research. The qual-

ity of these documents has crucial impact on the future access and comprehension of the archeological data from the excavations.

The purpose of archeological catalogues of artifacts is to support future research and reinterpretation of the data. They include several characteristics through which the artifacts are described. It is crucial that the descriptive terms are used in a uniform way; otherwise, the catalogue is impractical (Crook et al. 2002). The premise for future research is that, although cataloguing is implemented in a certain place and time, the terminology used should be applicable also in information retrieval and understanding (Taylor 2004). This is complicated by the heterogeneous nature of archeological data (Huvila 2006; Kintigh 2006a; Kintigh 2006b; Kintigh and Altschul, 2010; Snow et al. 2006; Richards 2002; Richards 2009; Richards and Hardman, 2008).

Discussion of the quality of archeological catalogues is hoped for, and they are required to be comprehensive, accurate, and logged in a format that is efficient and flexible to use (Crook et al. 2002). However, the nature and quality of cataloguing (i.e., as a process of knowledge organization) and catalogues (as a knowledge organization system) have rarely been the interest of archeologists, while in the field of knowledge organization (KO), this is an essential question (Hjørland 2008).

This paper combines the approaches of knowledge organization and archeology. It aims to analyse how archeological data is organized in the catalogues of archeological artifacts. Previously, Gnaden and Haldaway (2000) have studied observer-related variations in artifact recording and analyzed their effects statistically. Archeological cataloguing as classification and as a method for archeological analysis have been discussed by Dunnell (1986), Read (2009), and Rouse (1960), for example. In any case, these earlier studies have not focused on the aim of this article. In this study, archeological cataloguing is seen as a stage in which artifacts are described in digital format, as a database.

In general, knowledge organization systems (KOS) consist of classification and categorization schemes that organize content and control variant versions of key information. They may also include vocabularies, such as thesauri, semantic networks, and ontologies (Souza et al. 2012; Hodge 2000). To solve the problems of attainability and accessibility of cultural heritage data at global level, different kinds of knowledge organization systems have been developed for organizing of data. For example, the CIDOC-CRM (2011) is a standard conceptual reference model to combine and share data of museums, libraries, and archives and to map separate archeological records, glossaries, and thesauri with computer-based semantic web technologies (e.g., Doerr, Ore and Stead 2004; Doerr, Schaller and Theodoridou 2004). There exist established

vocabularies for artifacts and materials (e.g., *Material Thesaurus* 1997; *Object Thesaurus* 1999), and procedures for cataloguing such as *Artifact Cataloging System* (2007) and *Artifact Catalog Codes* (2012) for certain periods and groups of artifacts. There exist also efforts to combine vocabularies from different disciplines offered by web related technologies such as *Getty Vocabularies Web Services* (2013) that specialize in art, architecture and material culture. They offer vocabularies as a set of standard terminologies which are clustered with local vocabularies, e.g., local variant of terms, such as collection-specific terms, to support end-users (Harpring 2010). These vocabularies are, however, not applicable in all local collections.

Classification models for museum collections have been developed. A general taxonomy model of museum artifacts utilizes facets in classification: these are context facets (creator, style and period, geographical location), physical property facets (object type, material and technique), and motif facets (subject presented) (Ménard et al. 2010). The aim of this model is to organize data in a definite hierarchical structure and to offer a cognitive economy to increase effectiveness in use. However, it is essential to understand that archeological cataloguing for research and dissemination purposes differs from that of museums which have a non-research emphasis, i.e., exhibitions (e.g., Xia 2006) and different research emphasis. However, these goals are impossible to achieve without developing archeological cataloguing as a KOS.

Deokattey et al. (2010) describe vocabularies and classification schemas as conventional tools for developing knowledge organization, and ontologies as tools for representing and defining concepts and their relations systematically. A controlled vocabulary controls the variations of the use of synonyms and near-synonyms, homographs, and, e.g., grammatical variations by establishing a single form of the term (Noruzi 2006). Different kinds of controlled vocabularies such as classification schemes, thesauri, and taxonomies have had an import role in the organization and retrieval of information in different environments (Mai 2008). Both thesauri and ontologies are basically representational vocabularies for a specific domain, but the main difference is “the use of descriptors and concepts to map a given domain” (Deokattey et al. 2010, 174). Both controlled vocabularies and free index terms are used to create ontologies, but ontologies are more flexible than thesauri (Deokattey et al. 2010).

The main problem in culture heritage data sources, their interoperability in semantic level and publishing them as linked data (i.e., in their semantic aggregation) is their heterogeneity (Mäkelä et al. 2011). For example, reference vocabularies used for this purpose have not been mapped to each other. Ad hoc fixes in “inadequate databases” are created during cataloguing to describe objects,

they can differ from general semantics used in the discipline, and one item in a database can describe several objects creating compound textual explanations. (Mäkelä et al. 2011, 1, 3-8).

Knowledge organization has its focus on the nature and the quality of knowledge organizing processes and knowledge organizing systems which are used to organize documents, document representations, works, and concepts. The research has focused on studying activities such as document description, indexing, and classification conducted in memory institutions by different kinds of information specialists and computer based retrieval technologies (Hjørland 2008). Knowledge organization combines theoretical and practical contributions from different scientific disciplines to design systems to process information (Gonzalez 2007 cited in Bonome 2012). They consist of three main elements: knowledge, human beings, and automated systems, which are in dynamic and complex interaction (Bonome 2012). The main interest is in the design of efficient processes for the wide-scale knowledge representing, processing, and sharing (Bonome 2012) and promoting the retrievability of information (Souza et al. 2012).

Organizing knowledge includes three basic elements: the object that carries the content, i.e., knowledge, the context constituting the frame of reference, and the aim which is supported by organizing knowledge (Bonome 2012). Knowledge can be understood as a representation of information assimilated by a person, scientific or subjective, interpretative knowledge. The environment in which knowledge is generated can be described as context represented usually by organizations which influence human decision making. (Bonome 2012). In this case, the object is an archeological artifact, the context an archeological excavation, and reporting, including the cataloguing process, aiming to contribute archeological research in the future. The expected reliable contribution of an archeological catalogue in future research sets requirements for the archeological catalogue. This is one specific process in creating KOS for archeology.

There are no universal criteria or schema for archeological cataloguing, and, with regard to this study, no exact instructions for the cataloguing of artifacts from historical eras. The National Board of Antiquities (NBA) gives orders how to produce archeological reports including archeological catalogue of artifacts in Finland. The structure of an artifact catalogue is required to be systematic and to use correct terminology (Museovirasto 2010). However, there exist no local archeological vocabularies and cataloguing tools for Finnish archeologists, and free describing or indexing are the most-used methods. The parallel *Ontology for Museum Domain* (2011) does not fulfill the requirements of archeology. The current Museum 2015 project aims to create an architecture for

museums' collection management including guidelines for cataloguing archeological data resources using (Museovirasto 2013; NBA 2013) Spectrum 4.0 standard (KDK 2011; KDK 2012), indicating that there is an immediate need to research the quality of archeological cataloguing.

The cataloguing and description in archeology to some degree resemble that in libraries. They share the same purpose, to support the future use of the items. The difference between them is that in information science this procedure is called cataloguing (i.e., making a catalogue) and indexing (i.e., describing the content of the publication), whereas in general, in archeology, the previous term covers both functions. In the next section, we have a look at factors affecting the quality of cataloguing.

2.0 Factors affecting the quality of cataloguing

The main purpose of indexing is to describe the content of an item in a form suitable for inclusion in some type of database, from which it can be found on the basis of its content when needed in future. (Lancaster 1991). Indexing of the content consists of two phases, (1) conceptual analysis and (2) translation. Conceptual analysis includes the decision of what an item is about. Often this is done from the viewpoint of the interests of a future user audience. Translation means the description of the result of the conceptual analysis through index terms originating from an indexer's head or a controlled vocabulary. (Lancaster 1991). Lancaster (1991) refers to indexing of published and printed documents, but simultaneously also to the indexing of non-print documents, such as audio-visual, visual and sound media, and realia. Realia refers to virtual or other objects serving as illustration (Smith 1997), e.g., to archeological artifacts which could be visualized by using images or video clips.

The following factors, based on Lancaster (1991) may have an effect on the quality of indexing. Indexer factors, i.e., characteristics of the person conducting the indexing, include subject knowledge, experience, concentration, comprehension of the item, and knowledge of user needs. Vocabulary factors refer to the characteristics of the vocabulary used in indexing, and include specificity, ambiguity, or imprecision, quality of entry vocabulary, quality of structure, and availability of related aids. Document factors, in the case of archeology, refer to difficulties in the interpretation of the item. Process factors include type of indexing, rules and instructions, haste originating from required productivity and exhaustivity of indexing. Environmental factors include, for example, heating or cooling, lightning, and noise in the environment where indexing is conducted (Lancaster 1991).

An indexing failure may originate from misinterpretation of an item in question, from a difficulty in choosing

the most specific term to represent it, or from the use of an inappropriate term (Lancaster 1991.) In Lancaster's (1991, 76) words, indexers could be expected to "perform more effectively when they are given precise rules and instructions". In the field of archeology, the absence of standard guidelines causes problems (Crook et al. 2002.) Random errors are caused by errors in judgement of fluctuations in observation conditions. Systematic errors occur when an incorrect attribute is consistently applied to a class of artifacts. Illegitimate errors are genuine, accidental mistakes such as transcription errors (Gnaden and Holdaway, 2000; see also Crook et al. 2002).

3.0 Research questions

The aim of this article is to study what kind of data has been produced when cataloguing the archeological artifacts. The overall goal is to study the quality of an archeological catalogue of artifacts to understand how it could be developed when traditional human-based activities for information retrieval, integration and analysis of data are becoming assisted or even replaced by computer-based technologies (Hjørland 2008). The content of the artifact catalogue is a database-like structure of information consisting of stable columns for variables and rows for documented artifacts called subnumbers (Figure 1). Subnumbers or sub-units represent artifacts documented from the same context (i.e., strata where the artifacts were found) of the grid coordinate system in the excavation.

The research questions in this study are:

1. Which general concept variables are distinguished from the catalogue of archeological artifacts? Into which variables have the characteristics of the subnumbers been divided in the column for free description?
2. What is the quantitative distribution of these variables?
3. Is the cataloguing done systematically? What kinds of properties of the artifacts are catalogued and not catalogued? Is there any possibility to define the concepts not catalogued?

4. How do the characteristics in the free description column depend on other information in the artifact catalogue?; and,
5. How should archeological cataloguing be developed?

4.0 Material and methods

The research material of this study consists of urban archeological material from Lassfolk, Jakobstad (in Finnish Pietarsaari) in Finland. The site and its archeological layers were threatened by a construction project, and, in 2007 and 2008, the Lassfolk area was a research object of archeological excavations organized by the Department of Monuments and Sites, NBA. (Oikarinen 2008; 2009.) Excavation area C was chosen for the analysis because it included the most of artifacts with free description and it was the most rich in artifacts covering 910 subnumbers of artifacts. The whole excavation was much larger, but this amount is seen enough for analysis to produce qualitative results about the content.

The method applied in the study is content analysis and, more precisely, the content decomposition method (Tuomi and Sarajärvi 2009). Content analysis is defined as "distinguishing characteristics" or as "a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (Berelson 1971, 18). Content analysis means qualitative analysis whereas decomposition means the quantitative decomposition of content (Tuomi and Sarajärvi 2009). In this study, the more precise term "content decomposition method" is used, although qualitative analysis is used to define and classify the variables in the catalogue of artifacts. The quantitative analysis was conducted, applying SPSS statistical package to generate a general view of the results of the qualitative classification and description process. So far, the method has not been applied to archeological catalogues.

In the catalogue, variables for stable columns were main number (i.e., identification number for archeological artifact collection in certain excavation), subnumber (i.e., is different for each row in catalogue), coordinates (x, y, z, i.e., location), context (i.e., stratigraphic context of the ar-

Main number	Sub-number	Unit (context)	X-coordinate	Y-coordinate	Z-coordinate	Main material	Material	Type (item)	Description	Total	Measure (cm)	Weight(g)	Else
KM2008051	13	CSY2	478	476		Glass	Bottle glass		1 green, 1 transparent	2		11,7	
KM2008051	14	CSY2	478	476		Glass	Window glass		Green	1		1,4	
KM2008051	15	CSY2	478	476		Metal	Iron	Key	Big	1	10,9x3,2x0,9	46,2	
KM2008051	16	CSY3	488	476		Pottery	Stone ware	Vessel	Brownish glazing	1		20	

Figure 1. A part of a catalogue of artifacts translated from Finnish (Oikarinen 2009, Appendix 8).

tifant as codes), main material, material, type (i.e., artifact type), total (amount of artifact pieces), weight (as grams), and measures (verbal or numerical), as well as description for free expression and a column else (i.e., notes). A subnumber means one row, i.e., unit in the catalogue. The logic to catalogue the artifacts from one excavation area is to list them according to their material alphabetically, and in the order of their context number and x- and y-coordinates. (Figure 1.) One subnumber can consist of one or numerous pieces of artifacts. Consequently, the sample is enough to permit making general assumptions about cataloguing practises by methods used in analysis.

By using decomposition analysis, any document can be divided into different variables and values of variables, which can be studied and counted. The examined catalogue included artifacts described by the following variables: color (for example red), use (for example, a precise concept like knife from which it is also possible to conclude its function), surface finish (for example glazed), form (for example, flat), number (adjustment—for example, one handle), dating, size (in words like “small”), and burnt. In addition to these, there were variables, descriptionElse and datingElse. In the research material of this

study, the variables were coded as numbers: color relating to the number of colors mentioned (1 = no color mentioned, and 2 = one color mentioned etc.). The number of forms was coded through the same kind of system (1 = no form mentioned, etc.). The use, surface finish, number, dating, size, and burnt were coded in values according to mentions of them, like 1 = yes or 2 = no. The number of properties was counted as a sum for each subnumber as 1 = not mentioned, 2 = one property mentioned, etc. In addition, there was an option to create more variables and assign them more values, for example, in form. The data was described by statistical tables and diagrams.

5.0 Results

The archeological artifacts from the city of Jakobstad are typical urban artifacts from Finland dating from the 17th to the 19th century. Their most general main material group was pottery (49.3% of subnumbers), then glass (29.5%) and metal (16.8%). There were also small quantities of bones, leather, bark and stone, wood artifacts, and wool (Figure 2). The interest in this study was focused on free expression in the description column, which includes

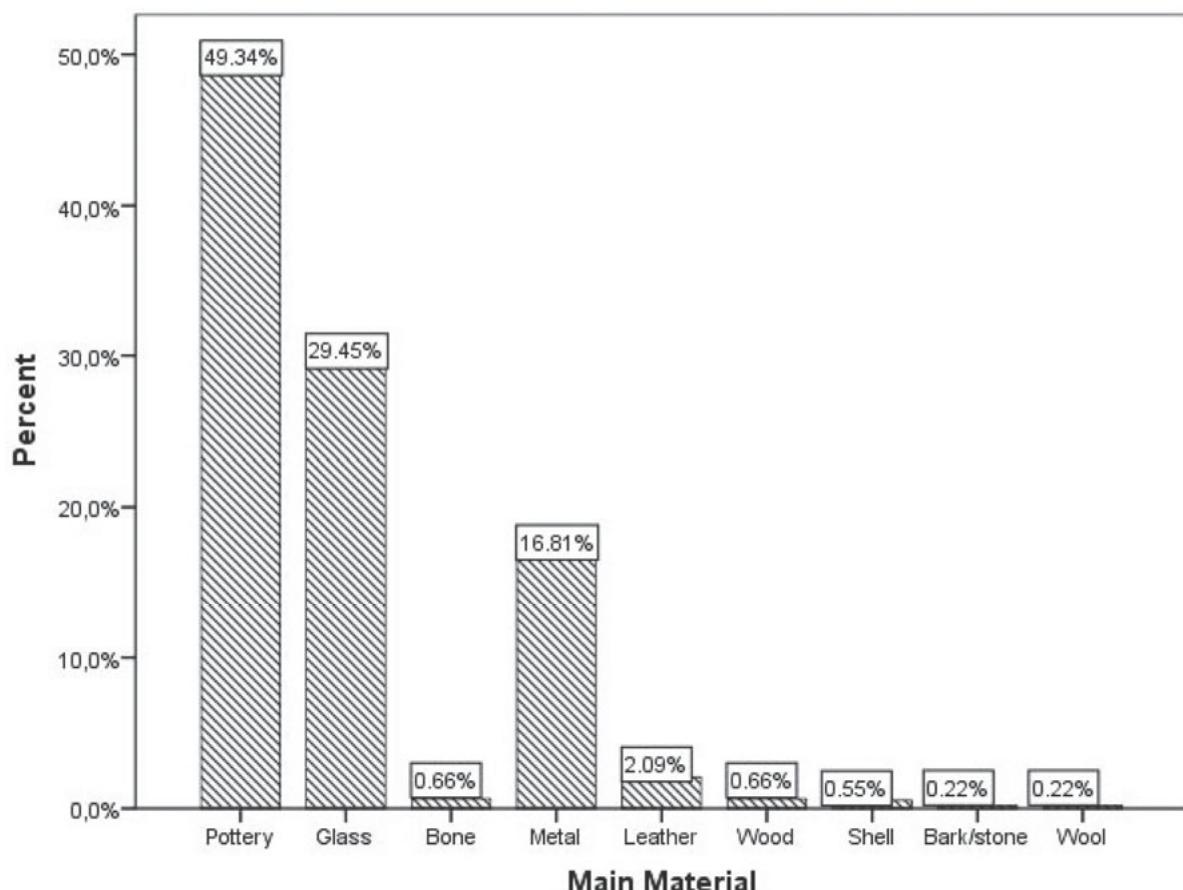


Figure 2. Relative proportions of subnumbers in percentages in the main material.

the following variables: color, use, form, surface finish, burnt, dating, and size. These variables were compared to information in other columns through cross-tabulation.

The main materials of the subnumbers were cross-tabulated with the characteristics describing them. Nine hundred ten subnumbers had received a total of 1,496 descriptions. A substantial number (82.3%) of the subnumbers had been described (Table 1). Two hundred sixty-five glass artifacts had received 268 descriptions (98.9%), for leather, the figures were respectively 18/19 (94.7%), and ceramics respectively, 377/499 (83.9%). This indicates that glass and ceramics are easy to describe or identify. Glass occurs in different colours and ceramics have a lot of characteristics to describe such as decoration or glaze, and moreover recognizable pieces, such as a piece of a handle.

Table 1 indicates that 17.7% of all subnumbers had not been described. The lowest percent of described subnumbers was in metals. About half of the metal subnumbers had been described, 73/153 (47.7%). An obvious reason

for this is that in the cataloguing phase they had not been restored and were unidentifiable. In this sample, all wool, bark and stone, wood, and bone artifacts had been described. However, not much can be concluded on the basis of these very sporadic (2-6) subnumbers (Table 1).

The descriptions of the subnumbers by classifying variables in the free description column were studied through cross-tabulation to see how many variables were used to describe each subnumber. The differences were clear: 17.7% of subnumbers had received no classifying variables, 37.9% had received one, and 21.5% two classifying variables. Some subnumbers were described by three (11.5%) or four (10.9%) variables. There were only a few subnumbers in which five (0.22%) or seven (0.11%) properties were mentioned (Figure 3). This result refers to non-uniformity and omission problems in indexing (Bernier 1980) but also to the diversity of archeological material.

The most common defined attribute in this research material is color, although it was not mentioned in 50.3%

Main material	The number of variables describing the sub-numbers							Sub-numbers total	Nr of described sub-numbers (1-7)	% of described sub-numbers
	0	1	2	3	4	5	7			
Ceramics	72	106	108	65	95	2	1	449	377	83.9
Glass	3	171	59	31	4	0	0	268	265	98.9
Bone	0	6	0	0	0	0	0	6	6	100
Metal	80	48	20	4	1	0	0	153	73	47.7
Leather	1	8	6	4	0	0	0	19	18	94.7
Wood	0	2	3	1	0	0	0	6	6	100
Shell	5	0	0	0	0	0	0	5	0	0
Bark and stone	0	2	0	0	0	0	0	2	2	100
Wool	0	2	0	0	0	0	0	2	2	100
Subnumbers total	161	345	196	105	100	2	1	910	749	82.3

Table 1. Totals of subnumbers in different materials and the number of characteristics describing them.

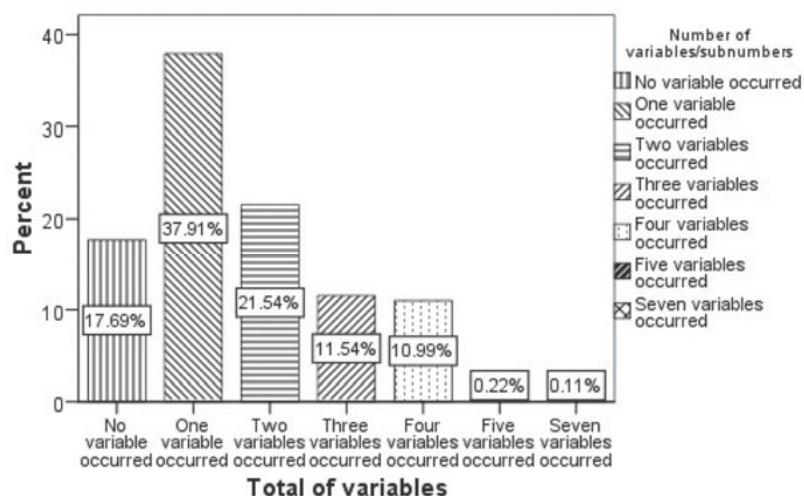


Figure 3. Relative proportions of subnumbers according to the number of variables applied in the free description of subnumbers.

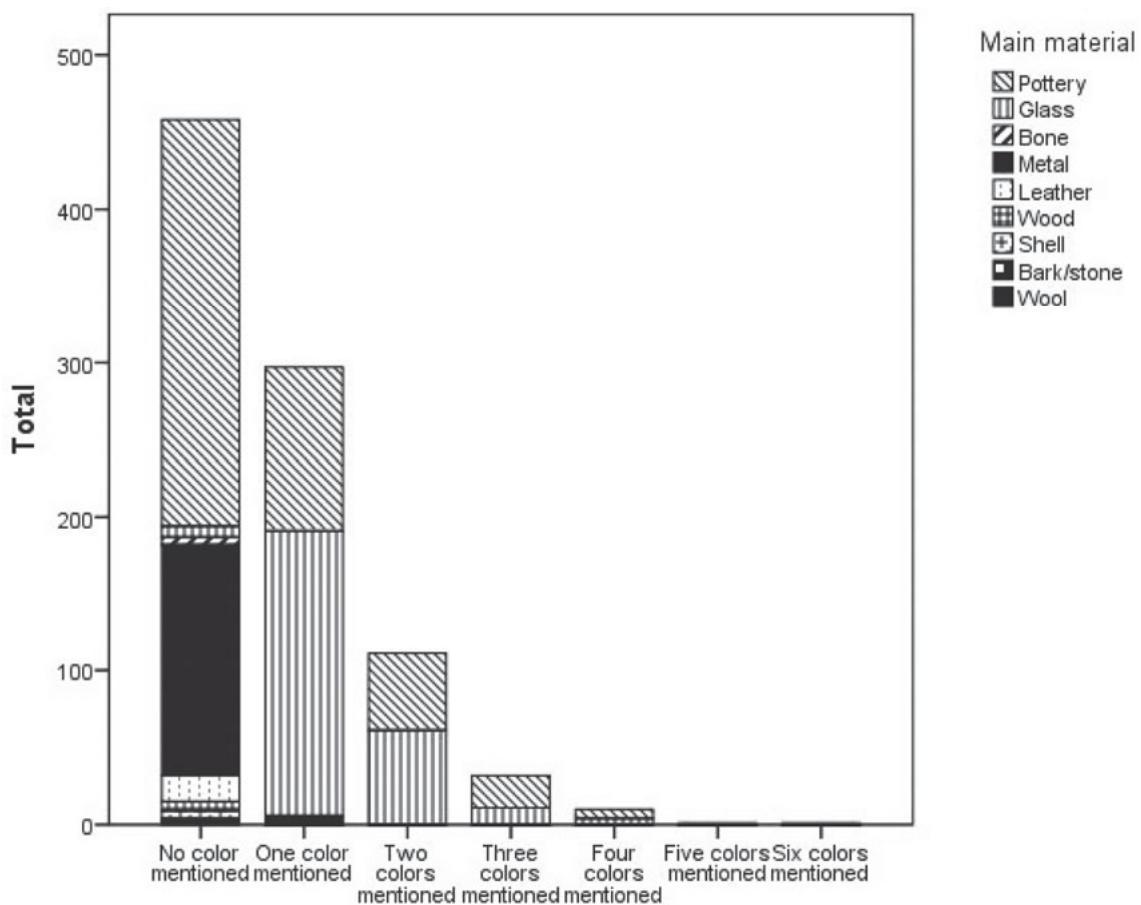


Figure 4. Number of colors mentioned per subnumber according to main materials.

of the subnumbers (Figure 4 and Table 2). Most often, there was a description of one (32.6%) or two (12.6%) colors. The most colorful materials according to the results are ceramic and glass. This outcome can be discussed, what it indicates, because all the artifacts have some kinds of colors. This result, too, indicates non-uniformity and omission problems in the indexing of the artifacts.

Color is a very subjective quality and easy to describe, but, without standards, it is very difficult to distinguish between different colors or shades. The most used color standard is the Munsell Colour Chart (Munsell 1907), but using it is very time-consuming and interpretative in nature (Goodwin 2000). It is debatable if the subjective color definition is informative for the user of the catalogue. When the color is typical for material and obvious for the cataloguer, it is not mentioned in the free description, such as red or white in ceramics. This is an established practice and a time-efficient way to catalogue when using established names for artifacts and their materials. Usually only special colors were mentioned, or typical but totally different variations in the material, such as green or transparent window glass. This is a tradition, and it optimizes consumption of time not to catalogue redundant

information. However, it produces non-uniformity, as all subnumbers are not treated in an equal way. It is debatable if the information is needed if the type of an artifact has been defined correctly such as red ware. Another option would be a column for color also to trace after interpreted colors for materials or to instruct more precisely what kind of other colors should be described.

Number of colors mentioned	Frequency of subnumbers	%
0	458	50.3
1	297	32.6
2	111	12.2
3	32	3.5
4	10	1.1
5	1	0.1
6	1	0.1
Total	910	100.0

Table 2. Totals of colors mentioned in the description column by proportions of subnumbers.

Colors are described most often in glasses (97.4% or 261/268 described subnumbers with 1-4 colors) and ceramics (41.2% or 185/449 subnumbers described with 1-

6 colors). This is mainly a representation of color of the material in glasses but decoration or glaze in ceramics. Also sporadic numbers of bone, metal, and leather were described with one color in this sample. Surface finishing was described or identified in 31.1% of the subnumbers of artifacts or pieces of them (Table 3). Most of them were ceramics (55.5% or 249/449 of subnumbers), bone (50% or 3/6 of subnumbers), wood artifacts (50% or 3/6 of subnumbers), and leather (36.8% or 7/19 of subnumbers) in this material. In almost 70% of the total amount of subnumbers, the surface finishing was not described. In general, there are many kinds of surface finishing in archaeological artifacts such as stitching in leather, or stamps, marks of manufacturing or decoration in glasses and pipes.

Variables described in subnumbers				
Yes	No			
Frequency of subnumbers	%	Frequency of subnumbers	%	Total of subnumbers
Use	365	40.1	545	59.9
Surface finish	283	31.1	627	68.9
Burnt	26	2.9	884	97.1
Dating	4	0.4	906	99.6
Size	12	1.3	898	98.7
				910

Table 3. Totals and relative proportions of the variables in the description column.

Number of forms mentioned	Frequency of subnumbers	%
0	841	92.4
1	32	3.5
2	20	2.2
3	15	1.6
4	2	0.2
Total	910	100.0

Table 4. Frequencies of the values of the variable form in the description column.

The concepts form and use are problematic variables to define when aiming at objectivity. It is practical to characterize the variable use as the original purpose of the item. Interpreting the life-time functionality as use of an archaeological artifact is not always straightforward (Bahn and Renfrew, 1997). The variable use was identified in 40.1% subnumbers (Table 3). This means that in about 60% of subnumbers, its function was not defined, whereas the classified variable form was not defined in 92.4% of subnumbers. (Table 4.) In many cases in the catalogue

of artifacts, there are words like piece, handle, or piece of bottom in ceramics, which are diagnostic properties and can be used to identify the use of an artifact, but they can also describe forms, or the amount of “pieces” of artifacts or that the artifacts are broken.

The results indicate confusion in the use of two concepts, material and type. For example, in the material column there was the concept “yarn,” even though it is type and should be recognized as material, i.e., as wool or something else. The confusion concerns especially fragile organic or metal artifacts which have to be collected with soil samples to preserve them after removal from their original context, and they need processing, which is done by a professional conservator. The cataloguing, however, is usually done before this phase. There are also materials and types which are rarer and difficult to recognize or define without conservation treatment and consulting with the specialists. This confusion refers to cases difficult to interpret and that, according to Lancaster (1991), belongs to problems caused by document factors. Moreover, the description column can contain information about use (i.e., indicating function and type). It can specify it or it can name it, even if the type, as a hyper concept has not been filled in. Sometimes the reader can deduce the use (i.e., a type) of the artifact by comparing the information contents in the type and description columns.

The content of type column (Table 5) shows that almost one third of subnumbers lack the definition of type (31.9%). There are mentions such as “an artifact” in the type column in 5.3% of artifacts, which refers to an unidentified type of artifact. Such ad hoc fixes during cataloguing are mentioned by Mäkelä et al. (2011). Most definitions of type refer to pottery originating from vessels (34.5%). These results imply that the vessels are easy to identify and cataloguers are familiar with them. Table 5 is a concrete example of a case where there exist no vocabularies to define the type and, consequently, the field is left empty or the terminology is sometimes haphazard.

	Frequency of subnumbers	%
Artifact type: not defined	290	31.9
Artifact type: vessel	314	34.5
Artifact type: vessel+?	1	0.1
Artifact type: artifact	48	5.3
Artifact type: miscellaneous artifact types	257	28,2
Total	910	100.0

Table 5. Distribution of subnumbers by artifact types mentioned in the type column.

The absence of information concerning material or type refers to problems in the interpretation of the artifact,

i.e., to document or indexer factors (Lancaster 1991, 79). Dating is hardly ever mentioned in the description column (0.4% of subnumbers) (Table 3).

The problems of non-uniformity in the description of the subnumbers are emphasized in the description of types, uses, forms, and colors of the artifacts. The analysis of uses, forms, and main materials showed that the variable form is mainly described for wood, leather, and metal. For example, from metal artifacts, it is difficult to recognize the type and, instead of this, there are adjectives describing them, which were recognized as forms. These refer to problems in interpretation and naming of the artifacts, i.e., document and vocabulary factors (Lancaster 1991). 64.4% of ceramic artifacts were described by their uses (289/449 of subnumbers). A large number of leather artifacts were described with terminology referring to their use (14/19 of subnumbers).

Occasionally (1.3% of subnumbers) in the description column, there was a verbal description of the size of an artifact, like “a big key” or measures of some special part of an artifact (Table 3). Also there were only sporadic mentions of burnt in subnumbers (2.9%), which merely means that burnt has been recognized in these cases (Table 3). The percentage of measured artifacts of all subnumbers is only 9.7% in the measure column of the artifact catalogue (Table 6). That means that, in this project, there was no time to measure more artifacts with informative properties and dimensions.

Measure	Frequency of sub-numbers		%
	Yes	No	
	88	822	90.3
Total	910		100.0

Table 6. Totals and relative proportions of values in the measure column.

6.0 Discussion

The creation of an archeological catalogue is a time-consuming and challenging phase after the field study, and the catalogue is needed as an access point to the artifact collection of the archeological site, also in the future. This emphasizes the need to study its quality.

The study focused on the description of archeological artifacts in the column for free description in the excavation catalogue. The most frequently occurring variables describing the studied 910 subnumbers were color, use, surface finish, and form, whereas burnt, dating, and size occurred infrequently. Although the information content in the catalogue was mainly systematic, the results reveal problems challenging its use and correct understanding in future research. Systematic errors were not identified, but

the catalogue contains random and illegitimate errors (Gnaden and Holdaway, 2000). All the factors which may have effect on the quality of content, i.e., indexer, vocabulary, document, process, and environmental factors (Lancaster 1991, 79) were recognized in this study.

The results imply to non-uniformity: the characteristics of some archeological artifacts had been described by several variables, while others were not described at all. Descriptions were mostly focused on inadequately identified artifacts. Artifacts of metal and leather, for instance, may be difficult to recognize before conservation treatment. Therefore, these may be described through subjective adjectives. Instead, in the cases of well-known artifact types, free descriptions either are sometimes missing, or they are very carefully catalogued depending on time schedule of the project. Redundant information, such as general color of the artifacts, is often left out of the catalogue. There are also mentions of details difficult to generalize as variables, for example, artifacts that are broken but described as well-preserved. In general the condition of artifacts is not reported.

The understanding of the columns material and type were sometimes mixed, or they had not been identified, or the data entry was empty. To identify uncommon artifacts as types, profound knowledge is needed or an excellent collection to compare them with. This appears also in the use of special vocabulary in the naming of manufacturing techniques or decoration in the description column. Facts, which are familiar to a cataloguer, are easy to name, such as different kinds of colors or surface finish. It would be possible to classify the latter into at least two more categories, namely manufacture and decoration. The same concerns the concepts use and form, which could be divided into more precise sub-variables. These variables were easily mixed implying that the terminology used in the catalogue is difficult to categorize.

Moreover, the catalogued information may be misleading in some cases: for example, the description column included a mention of three pieces of clay pipes with decoration, although besides them there were also three undecorated pieces. The total number of pieces (in this case six) is calculated in the total column. A reader has to be aware of such cataloguing practices to interpret the catalogue correctly.

All the practices and problems mentioned above reflect “constraints of a scholarly domain” originating from its discourse, history, schools of thought, paradigms, research fronts, activities, etc. (Mai 2008, 23). Reasons for the problems derive from conventions of archeological cataloguing, circumstances, and lack of appropriate tools. When cataloguing artifacts, archeologists still create classifications using generalized terms or terms created by themselves. This can be seen as subjective knowledge originating from cul-

ture, for example, or as assimilated knowledge that a person gives an expression (Bonomo 2012). Their non-standardized use makes information content incomparable between documents and collections, and difficult to search. In general, the purpose of the catalogue is to aid in identifying objects and providing access to artifact collections, which is enhanced by information retrieval by index term of free description retrieval.

Detailed descriptions may reveal errors in general level categorizations (Crook et al. 2002). Nevertheless, free expressions and comments may also include valuable archeological information. Therefore, there is a need to preserve unique local vocabulary and cultural history, for example, in the naming of the artifacts after their original names or atypical uses or reuses. The problem of heterogeneity related to archeological documents and collections is worldwide (Snow et al. 2006). In rescue excavations, there are usually no possibilities to correct the catalogue after the ending of the projects, although treatment and identification of artifacts still continue. Users do not know that the interpretation of artifacts is sometimes preliminary. These problems refer to indexer and process factors mentioned by Lancaster (1991).

These problems could be neutralized by using vocabularies to identify and name different materials and artifact types, and to describe them correctly. They could act as entries for index terms in information retrieval. A proper terminology in naming could contribute also in aggregating and linking (i.e., as cross-references) between different documents in digital formats. Also metadata relating the document should be standardized. If the access to actual artifact collection is not possible to gain, the most reliable way to use catalogued data is with digital images of artifacts.

Moreover, from a database-like structure, the artifacts are difficult to combine with for example the context descriptions in the report and other documented data concerning the same excavation although this is needed. For example, because the context of the artefacts (i.e., strata) is presented as codes, to understand the content of the context column other archeological documents are needed. If the catalogue has not been created systematically, the amounts of different kinds of materials, artifact types, their weight and division in documented contexts (i.e., in strata type) etc. can not be analyzed statistically.

In archeology heterogeneous—or diverse—information is the fact that we have to accept (e.g., Snow et al. 1996), but new technologies may allow production of locally established content which is retrievable and also understandable at the global level. For example, new artificial intelligence and Semantic Web technologies such as natural language processing (NLP) have already been experimented to create ontologies in archeology, with promising results.

The achieved semantic indexing of archeological documents requires incorporation of existing ontologies and controlled vocabularies as conceptual framework (Vlachidis et al. 2010). At the same time, the usefulness and realizability of vocabularies have been discussed critically from the viewpoints of digital data preservation and integration related to terminological specialities and different languages (Eiteljorg 2011). However, controlled terminologies exposed in the Web, which are also applicable by and combinable in other programmatic solutions may be useful for creating knowledge organization systems, such as the recently proposed Terminology Web Services for archeology (Binding and Tudhope 2010).

Flexibility contributes to future classification (Bowker and Star 1999), and the challenge is to be aware of the effects of pre-established knowledge, technologies, and vocabularies in decisions made during cataloguing such as in artifact categorizations. Lock (2003, 82) has warned of this issue at general level: “The intentions of an analysis should not be determined by what the technology will do but by the archeological questions being asked.” This is important because archeological research is interpretative in nature. To be applied to archeological cataloguing, the components or vocabularies in the user interface should not restrict too much free description, but, on the other hand, should allow the analysis of the data for specific research questions. Also balancing the needs between global standardisation and local free description is needed.

The decomposition of archeological catalogue in this study produced new categorizations for archeological artifacts. These are: decoration, surface finish, marks of manufacturing, color, condition, dating, diagnostic properties (contributing to the identification of the artifact), and special properties or functions, such as reuse. These could be used together with the existing categorizations coordinates, i.e., location, context (i.e., stratigraphic context of the artifact), main material, material, type, (free) description, total (i.e., amount of pieces), measure (measured dimensions), weight and else (i.e., notes) in developing systems or meta-structure for documents.

As a result of this study, it is possible to extract recommendations: 1) vocabularies would improve the quality of the catalogue; 2) an effective cataloguing tool with guidelines, vocabularies, and a possibility to combine information from different types of digital documents and comparative archeological collections would improve the interpretation (i.e., research and analysis process); 3) the existence of combined (linked or integrated) archeological data resources such as artifact catalogues, images, maps, contextual and textual information, and/or access to them could help archeological research and analysis. According to literature, controlled vocabularies increase

possibilities to the use of latest technologies in making the disparate catalogues interoperable and usable in comparative archeological studies (Vlachidis et al. 2010). This study also revealed the need for localized tools and more specific guidelines in cataloguing.

There are several general requirements for an archeological catalogue in digital format: 1) it should describe the catalogued artifacts systematically and understandably to the user who usually is an archeologist; 2) it should preserve the local nuances of artifacts, and describe them at a consistent level; 3) it should work as a database which is analyzable for research purposes; and 4) it should be comparable and combinable with other relevant archeological data sets. The cataloguing should be flexible and controlled by cataloguers even if cataloguing tools would be available.

In this specific local context, the development also includes the challenge of unifying the goals of museum and archeological cataloguing principles as a coherent whole serving both disciplines (NBA 2013), although this is beyond the scope of this paper.

7.0 Conclusions

Keeping in mind the unique nature of archeological data and the need for unique descriptions, this study reveals some major problems in cataloguing large collections of archeological artifacts. The lack of standards and uniform practices in content descriptions results in subjectivity and incorrect or irrelevant vocabulary, which can lead to a danger of misinterpreting the content and to incomparability between documents and catalogues. This can also have an effect on the reliability of archeological field work reports. The study proves a need for localized vocabulary developing towards ontologies in archeology, and contributes to the development of tools and practices in the cataloguing of archeological artifacts and their future use in digital environments.

Due to the increasing possibilities and utilization of Web-related technologies and information technologies today, knowledge organization systems may extend from local to almost global scale, and they spread over the capacity of organizations to set standards or rules to be implemented in them. The Web increases possibilities for interaction and has an impact on knowledge creation by offering new channels for interaction and for creating social networks, new communication tools, and ways how to process and retrieve information. It blurs traditional organizational boundaries as a context of interaction and self-recognition (Bonomo 2012).

These changes also concern the setting of requirements for future archeological knowledge organization systems, and the process of archeological cataloguing. This study revealed problems mainly in the naming and describing of

archeological artifacts. This refers to broader issues in conceptualizations and reporting in archeological discipline reflecting also to the heterogeneous nature of archeological data (e.g., Huivila 2006; Snow et al. 2006).

Archeological catalogues reflect Xia's view (2006, 271) of the difficulty of using "patterned descriptions to precisely elucidate variations of individual objects." This is solved by using individual descriptions ending up in non-systemicity in the catalogues. Solution to this and the enormous quantity of archeological data could be electronic publishing (Xia 2006) combined with shared vision of knowledge organization systems discussed by for example Bonome (2012), Souza et al. (2012), and da Silva and Ribeiro (2012).

At the same time, the introduction of user-generated free-form tags, or folksonomies seems to be removing hierarchy from the scheme of knowledge organization through facilitating knowledge discovery and web indexing (Noruzi 2006), which could also contribute to content description or creating metadata for archeological data. This is an example of a KOS which integrates knowledge contributed by individuals from the social networks (Bonomo 2012), e.g., of archeologists and other relevant specialists.

Further requirements are set by Mai (2008) who calls for a more domain-centered approach in the design of controlled vocabularies where knowledge and expertise of indexing should be complemented with that of information behavior to match the actors' information needs. A comprehensive archeological KOS should be based on the cognitive work analysis of archeologists (Mai 2008), which combines the cataloguing of the artifacts with their context information and all the other information (i.e., varying data resources) produced in the excavation.

References

- Artefact catalog codes—The London Museum of Archaeology.* 2012. London Chapter, Ontario Archaeological Society. Available <http://www.ssc.uwo.ca/assoc/oas/misc/catcodes.html>.
- Artifact cataloging system.* 2007. Society for Historical Archaeology (SHA). The University of Montana. Available http://www.sha.org/research/artifact_cataloging_system.cfm.
- Bahn, Paul and Renfrew, Colin. 1996. *Archaeology: theories, methods and practice*. London: Thames and Hudson Ltd.
- Berelson, Bernard. 1971. *Content analysis in communication research*. New York: Hafner Publishing Company.
- Bernier, Charles L. 1980. Subject indexes. In Kent, Allen, Lancour, Harold and Daily, Jay E., eds., *Encyclopedia of library and information science*, 29. New York: Marcel Dekker Inc, pp. 191-205.

Binding, Ceri and Tudhope, Douglas. 2010. Terminology web services. *Knowledge organization* 37: 280-6.

Bonomé, María G. 2012. Analysis of knowledge organization systems as complex systems: A new approach to deal with changes in the web. *Knowledge organization* 39: 104-9.

Bowker, Geoffrey C. and Star Leigh, Susan. 1999. *Sorting things out: Classification and its consequences*. Cambridge (Mass.); London: MIT Press.

CIDOC-CRM. 2011. CIDOC Conceptual Reference Model. International Council for Museums. Available <http://www.cidoc-crm.org/>.

Crook, Penny, Lawrence, Susan and Gibbs, Martin. 2002. The role of artefact catalogues in Australian historical archaeology: A framework for discussion. *Australasian historical archaeology* 20: 26-38. Available http://ashadocs.org/aha/20/20_04_Crook.pdf

Deokattee, Sangeeta, Neelameghan, Arashanipalai and Kumar, Vijai. 2010. A method for developing domain ontology: A case study for a multidisciplinary subject. *Knowledge organization* 37: 174-84.

Doerr, Martin, Ore, Christian-Emil and Stead, Stephen. 2004. The CIDOC reference model – A New standard for knowledge sharing. In J. Grundy, John C. Grundy, Hartmann, Sven, Laender, Alberto H.E., Maciaszek, Leszek A. and Roddick John F, eds., *Proceedings, tutorials, posters, panels and industrial contributions at the 26th International Conference on Conceptual Modeling - ER 2007 Auckland, New Zealand. CRPIT, 83*, pp. 51-56. Available <http://crpit.com/confpapers/CRPITV83Doerr.pdf>.

Doerr, Martin, Schaller, Kurt and Theodoridou, Maria. 2004b. Integration of complementary archaeological sources. In Niccolucci, Franco and Hermon, Sorin, eds., *Proceedings of the International Conference on Computer Applications and Quantitative Methods in Archaeology Conference, CAA 2004, 13-17 April, 2004, Prato, Italy*. Budapest: Archeolingua. Available http://www.ics.forth.gr/isl/publications/paperlink/doerr3_caa2004.pdf.

Dunnell, Robert C. 1986. Methodological issues in Americanist artifact classification. *Advances in archaeological method and theory* 9: 149-207.

Eteljorg, Harrison. 2011. What are our critical data-preservation needs? In Kansa, Eric C. et al. eds., *Archaeology 2.0. new approaches to communication and collaboration*. Los Angeles: Cotsen Institute of Archaeology Press, pp. 251–64. Available <http://www.escholarship.org/uc/item/1r6137tb>.

Getty Vocabularies. 2013. Getty Research Institute. J. Paul Getty Trust. Available <http://www.getty.edu/research/tools/vocabularies/index.html>.

Gnaden, Denis and Holdaway, Simon. 2000. Understanding observer variation when recording stone artifacts. *American antiquity* 65: 739-47.

Goodwin, Charles. 2000. Practices of color classification. *Mind, culture, and activity* 7: 19-36. Available http://www.sscnet.ucla.edu/clic/cgoodwin/96col_class.pdf

Harpring, Patricia. 2010. *Introduction to controlled vocabularies. Featuring the Getty vocabularies*. Available http://www.getty.edu/research/tools/vocabularies/intro_to_vocabs.pdf

Hjørland, Birger. 2008. What is knowledge organization (KO)? *Knowledge organization* 35: 86-101.

Hodge, Gail. 2000. *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Washington D.C: The Digital Library Federation Council on Library and Information Resources. Available <http://www.clir.org/pubs/reports/pub91/reports/pub91/pub91.pdf>

Huvila, Isto. 2006. *The ecology of information work: A case study of bridging archaeological work and virtual reality based knowledge organization*. Åbo: Åbo Akademis förlag - Åbo Akademi University Press.

Kansallinen digitaalinen kirjasto (KDK) [National Digital Library]. 2011. *Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri: Liite B standardisalkku*. Kansallinen digitaalinen kirjasto. Available <http://www.kdk.fi/images/stories/tiedostot/kdk%20standardisalkku%202011-08-29.pdf>.

Kansallinen digitaalinen kirjasto (KDK) [National Digital Library]. 2012. *Kansallinen digitaalinen kirjasto –hanke. Välimääräajalta huhtikuu 2011 – kesäkuu 2012*. Kansallinen digitaalinen kirjasto. Available http://www.kdk.fi/images/stories/tiedostot/KDK-vliraportti_27092012-1.pdf.

Kintigh, Keith. 2006a. The challenge of archaeological data integration. In *Union Internationale des Sciences Préhistoriques et Protohistoriques, Lisbon, 4-9 September 2006*. Available <http://archaeoinformatics.org/articles/Kintigh2006UISPP.pdf>.

Kintigh, Keith. 2006b. The promise and challenge of archaeological data integration. *American antiquity* 71: 567-78. Available <http://www.public.asu.edu/~kintigh/Kintigh2006CyberinfrastructureAmAnt.pdf>

Kintigh, Keith and Altschul, Jeffrey H. 2010. Sustaining the digital archaeological record. *Heritage management* 3: 264-74. Available <http://www.digitalantiquity.org/wp-uploads/2011/01/20110127-Kintigh-Altschul-Forum-on-Sustaining-the-Digital-Archaeological-Record-from-Heritage-Management.pdf>

Lancaster, Frederick Wilfrid. 1991. *Indexing and abstracting theory and practise*. Champaign, Ill: University of Illinois, Graduate School of Library and Information Science.

Lock, Gary. 2003. *Using computers in archaeology*. London; New York: Routledge.

Mäkelä, Eetu, Hyvönen, Eero and Ruotsalo, Tuukka. 2011. How to deal with massively heterogeneous cultural heritage data – lessons learned in CultureSampo.

Semantic web 0: 1-24 Available http://www.semantic-web-journal.net/sites/default/files/sw160_0.pdf.

Mai, Jens-Erik. 2008. Actors, domains, and constraints in the design and construction of controlled vocabularies. *Knowledge organization* 35: 16-29.

Material Thesaurus. 1997. *British Museum*. Trustees of the British Museum. Available http://www.collectionslink.org.uk/assets/thesaurus_bmm/matinintro.htm?phpMyAdmin=OYNyINPdn3sQmoXugKH1gcCLSW0.

Ménard, Elaine, Mas, Sabine and Alberts, Inge. 2010. Faceted classification for museum artefacts. *Aslib proceedings: New information perspective* 62: 523-32. Available <http://www.emeraldinsight.com/journals.htm?articleid=1876753&show=pdf>

Munsell, Albert H. 1907. *A color notation*. Boston: Geo. H. Ellis co. Available <http://www.gutenberg.org/etext/26054>

Museovirasto [National Board of Antiquities]. 2010. *Objekta historiallisen ajan kaivauuskertomuksen laatijalle. Rakennushistorian osasto*. Available <http://www.nba.fi/fi/File/1028/tutkimuskertomusohje-rho-2010.pdf>

Museovirasto [National Board of Antiquities]. 2013. *Museoiden kokoelmahallinnan kokonaisarkkitehtuuri 1.0. Museoviraston objekta 2. Museo 2015. Museovirasto*. Available <http://www.nba.fi/fi/File/1806/kokonaisarkkitehtuuri.pdf>

National Board of Antiquities [NBA]. 2013. The museum 2015 project. Available http://www.nba.fi/en/development/museum_2015

Noruzi, Alireza. 2006. Folksonomies: (Un)controlled vocabulary? *Knowledge organization* 33: 199-203.

Object Thesaurus. 1999. *British Museum*. Trustees of the British Museum. Available http://www.collectionslink.org.uk/assets/thesaurus_bmon/Objintro.htm?phpMyAdmin=OYNyINPdn3sQmoXugKH1gcCLSW0.

Oikarinen, Teija. 2008. *Pietarsaari, 2/2/7-10, Lassfolkin kortteli (PSL-07)*. Kaupunkiarkeologinen koekaivaus 6.9.-28.9.2007. (Unpublished excavation report) Museovirasto: Rakennushistorian osasto.

Oikarinen, Teija. 2009. *Pietarsaari, 2/2/7-10, Lassfolkin kortteli (PSL-08)*. Kaupunkiarkeologinen pelastuskaivaus 19.5.-8.8.2008. (Unpublished excavation report) Museovirasto: Rakennushistorian osasto.

Ontology for Museum Domain (MAO). 2011. Semantic Computing Research Group. Available <http://www.seco.tkk.fi/ontologies/mao/>.

Read, Dwight W. 2009. *Artifact classification: A conceptual and methodological approach*. Walnut Creek, Calif.: Left Coast Press.

Richards, Julian D. 2002. Digital preservation and access. *European journal of archaeology* 5: 343-66. Available <http://eja.sagepub.com/cgi/content/abstract/5/3/343>.

Richards, Julian D. 2009. From anarchy to good practice: The evolution of standards in archaeological computing. *Archeologia e calcolatori* 20: 27-35. Available http://eprints.whiterose.ac.uk/10707/1/Arch_%26_Calcolatori_standards_2009_richards.pdf.

Richards, Julian D. and Hardman, Catherine. 2008. Stepping back from the trench Eege. An archaeological perspective on the development of standards for recording and publication. In Greengrass, Mark and Hughes, Lorna eds., *The virtual representation of the past*. Farnham, Surrey and Burlington, USA: Ashgate Publishing Company, pp. 101-12. Available <http://eprints.whiterose.ac.uk/7795/1/richardsjd4.pdf>

Rouse, Irving. 1960. The classification of artifacts in archaeology. *American antiquity* 25: 313-23.

da Silva, Armando Malheiro and Ribeiro, Fernanda. 2012. Documentation / information and their paradigms: Characterization and importance in research, education and professional practise. *Knowledge organization* 39: 111-24.

Smith, Bryan. 1997. Virtual realia. *The Internet TESL journal* III (7). Available <http://iteslj.org/Articles/Smith-Realia.html>.

Snow, Dean. R., Gahegan, Mark, Giles, C. Lee, Hirth, Kenneth G., Milner, George R., Mitra, Prasenjit and Wang, James Z. 2006. Cybertools and archaeology. *Science* 311: 958-9. Available http://clgiles.ist.psu.edu/papers/Science-cyber-historical_2006.pdf.

Souza, Renato Rocha, Tudhope, Douglas and Almeida, Maurício Barcellos. 2012. Towards a taxonomy of KOS: Dimensions for classifying knowledge organization systems. *Knowledge organization* 39: 179-92.

Taylor, Arlene G. 2004. *The organization of information*. Westport, Connecticut; London: Libraries Unlimited.

Tuomi, Jouni and Sarajärvi, Anneli. 2009. *Laadullinen tutkimus ja sisällönanalyysi*. Helsinki: Kustannusosakeyhtiö Tammi.

Vlachidis, Andreas, Binding, Ceri, Tudhope, Douglas and May, Keith. 2010. Excavating grey literature: A case study on the rich indexing of archaeological documents via natural language-processing techniques and knowledge-based resources. *Aslib proceedings* 62: 466-75.

Xia, Jingfeng. 2006. Electronic publishing in archaeology. *Journal of scholarly publishing* 37: 270-87. Available <http://arizona.openrepository.com/arizona/bitstream/10150/105391/1/Archaeology.pdf>