Marta Kuźma*

Are You Able to Find the Maps You Need?

Keywords: map metadata, raster map, metadata evaluation.

Summary: Currently, searching for objects in digital libraries is based on metadata. That is why the description of cartographic documents in academic and national libraries was assessed. The ways of describing map metadata and the standards being the basis for description were analyzed. Finally, a statistical analysis was carried out, which allowed determining in which library the metadata is best developed.

Introduction

Access to archived maps in digital form is becoming simpler (Moore 2005). Maps are collected by archives, libraries and made available digitally through dedicated portals. Information about archival cartographic resources is also disseminated by such content aggregators as Europeana (Europeana 2019), Digital Public Library of America (DPLA 2019), and The Institute of Museum and Library Services Digital Collections and Content (IMLS DCC 2019). Despite so many different map access options, it is not easy to find the maps you need. The main users of archival maps include geographers, historians, sociologists, urban planners, and others. They use these maps for scientific, historical, and geographical research, as well as for didactic purposes (Youngblood 2006). Searching for objects in digital libraries is based on metadata. Therefore, in order to enable specific users to find archival maps that suit their needs, research was conducted on the quality of archive map metadata in digital libraries. For the purposes of this study, metadata quality is understood as completeness and consistency. Completeness means the degree to which objects are described using all possible metadata elements, and consistency – that the same values represent similar concepts and the structure of metadata representation is similar, at a similar level of detail (Park, 2009), (Park and Tosaka, 2010).

Research Background

Methodology

The basis for the evaluation was the method of describing archival maps in digital libraries, the possibility of obtaining specific evaluation metadata, and the completeness of archival map metadata. The metadata quality evaluation will allow the user to determine whether they can find maps to meet their needs. The metadata quality assessment was carried out as follows:

- Development of a set of archive maps metadata elements,

* Assistant Professor; The Military University of Technology Warsaw, Faculty of Civil Engineering and Geodesy [marta.kuzma@wat.edu.pl]

- Defining the rules for obtaining metadata from the digital library to the proposed set of metadata,
- Statistical evaluation showing which library best collects metadata of cartographic documents.

The first stage concerned the way in which archival maps should be described, i.e. the information that should be collected in metadata. As part of this stage, the standards used to describe objects in digital libraries – Dublin Core (DCMI 2012) and MARC21 (LoC MARC 2019) and the standard to describe geographical objects – International Organization for Standardization (ISO) 19115 (ISO 2014) were analyzed. Next, it was shown how metadata can be extracted from individual digital libraries into the proposed set of metadata elements. This allowed us to present the consistency of metadata in individual libraries. Next, the completeness of the metadata was checked by checking the number of objects with which the individual metadata elements were supplemented.

Data

The subject of quality evaluation was the metadata of archival maps collected in Polish digital libraries whose cartographic resources exceeded 100 objects. Four academic libraries (Jagiellonian Digital Library (JDL 2019), Digital Library of University of Wroclaw (DLUW 2019), Silesian University of Technology Digital Library (SUTDL 2019), Maria Curie-Skłodowska University Digital Library (MSCUDL 2019)), and the National Library of Poland (National Library of Poland 2019) were selected. They make their collections available through different channels to increase the number of users, for example the National Library of Poland publishes them through the Polona portal (Polona 2019), online catalogs (National Library of Poland 2019b) and the API service (National Library of Poland 2019a).

Table 1 presents the size of the cartographic resources in individual digital libraries. In academic libraries, all cartographic documents were analyzed: maps, atlases, and plans, while in the National Library of Poland – only topographic maps.

Item	Digital Library name	Number of archival cartographic documents
1	Jagiellonian Digital Library (JDL)	559
2	Digital Library of University of Wroclaw (DLUW)	1 733
3	Silesian University of Technology Digital Library (SUTDL)	107
4	Maria Curie-Skłodowska University Digital Library (MSCUDL)	239
5	National Library of Poland (NLP)	35 090

Table 1: Number of archival cartographic documents in each library.

The analyzed academic libraries collect metadata in the dLibra system, whose metadata scheme is based on the Dublin Core standard (DCMI 2012), while in The National Library of Poland the metadata is collected based on the MARC21 standard (LoC MARC 2019).

Set of metadata elements

The quality in this case will be related to a set of metadata elements of archival cartographic materials that determine how the maps fit the needs of users (Dimitrov 1998). Maps in digital libraries can be treated in at least two different ways: firstly, as library objects and secondly as geographical objects.

Therefore, the metadata set was developed based on the analysis of 3 different standards - two describing objects in digital libraries – Dublin Core and MARC21 and one describing metadata of geographical objects - ISO 19115. Finally, 13 elements of metadata were proposed (Table 2). The proposed set contains metadata that is typical of all objects in digital libraries (1-6) and characteristic for cartographic documents (7-13).

No.	Metadata elements	Kind of metadata element
1	Type of content	typical
2	Date	
3	Date range	
4	Main subject	
5	Rights	
6	Language	
7	Geographic location (geographic coordinates)	cartographic
8	Scale of map	
9	Reference system	
10	Mapping methods	
11	Map format	
12	Orientation	
13	Source materials used to develop the map	

Table 2: The proposed set of metadata elements.

The recording of metadata in digital libraries is not always standardized, which means that every librarian can fill up the metadata according to their own experience. In order to obtain metadata from digital libraries for the proposed set of metadata elements (Table 1), the methods of assigning metadata were defined and weighed as presented in Table 3. (Kuźma, Mościcka 2018).

No.	Features	Weight	Description
1	Directly	1.0	Data (values stored in individual metadata elements) taken from a particular metadata element are directly attributed to the corresponding evaluation criterion, e.g., the language of cartographic resources (metadata in digital library) corresponds to the language (metadata element in the proposed set of metadata elements).
2	Simple analysis	0.8	Data for several evaluation criteria are derived from one metadata element. These metadata records are clearly separated from each other (e.g. by a semicolon or comma). These data are divided and attributed to appropriate evaluation criteria, such as "Date range" and "Geographic location" (evaluation criteria) that are derived from a single "keyword" element (metadata). The data in the original metadata are separated by a semicolon
3	Specialist analysis	0.5	Metadata values are derived from metadata elements based on the operator's experience and attributed to the relevant evaluation criteria
4	No data	0.0	Data not available

Table 3: Features and weights of obtaining metadata.

The level of difficulty of obtaining data

The summation of weights for a typical metadata element in a particular library made it possible to determine the level of difficulty of obtaining data for a typical (E_{1l}) metadata element in each of the libraries:

$$E_{1l} = \sum_{e=1}^{6} w_{le} \quad (1)$$

The summation of weights for a cartographic metadata element in a particular library made it possible to determine the level of difficulty of obtaining data for a cartographic (E_{2l}) metadata element in each of the libraries:

$$E_{2l} = \sum_{e=7}^{13} w_{le} \quad (2)$$

Where:

l – Digital library number (Tab. 1. Item),

e – Evaluation element number (Tab. 2. No.),

w_{le} – weight value for number e evaluation element in number l digital library

The following scale was used to assess the level of difficulty of obtaining data for typical metadata elements (E_{1l}) in individual libraries:

- 0 - 3.0 – poor because there were no data or data were derived from a specialist analysis (the upper limit was taken as the weight of 0.5x6 metadata elements);
- 3.1 - 4.8 – sufficient because the data were obtained at least on the basis of simple or specialist analysis;
- 4.9 - 5.4 – good, because data for at least half of the criteria were obtained by using at least a simple analysis (the lower limit of the range was assumed to be the weight of 0.8x6 metadata elements);
- 5.5 - 6.0 – very good, because the data for at least half of the metadata elements were obtained directly, and for the rest of the elements by simple analysis (the lower limit of the range was assumed to be: the weight of 1x3 metadata elements plus the weight of 0.8x3 metadata elements divided by two).

The following scale was used to assess the level of difficulty of obtaining data for typical metadata elements (E_{2l}) in individual libraries:

- 0 - 3.5 – poor because there were no data or data were derived from a specialist analysis (the upper limit was taken as the weight of 0.5x7 metadata elements);
- 3.6 - 5.6 – sufficient because the data were obtained at least on the basis of simple or specialist analysis;

- 5.7 - 6.2 – good, because data for at least half of the criteria were obtained by using at least a simple analysis (the lower limit of the range was assumed to be the weight of 0.8x7 metadata elements)
- 6.2 - 7.0 – very good, because the data for at least half of the metadata elements were obtained directly and for the rest of the elements by simple analysis (the lower limit of the range was assumed to be the weight of 1x3 metadata elements plus the weight of 0.8x4 metadata elements divided by two).

The summation of weights from all digital libraries for each specific metadata element made it possible to determine the level of difficulty of obtaining data in a digital library for each metadata element (E_{1e}):

$$E_{1e} = \sum_{l=1}^{5} w_{le} \tag{2}$$

The following scale was used to evaluate the level of difficulty of obtaining data in a digital library for each metadata element (E_{1e}):

- 0-2.5 – poor, because the data for the criterion were obtained based on specialist analysis (upper limit of the range: weight of 0.5x5 libraries);
- 2.6-3.9 – sufficient because the data for the criterion were obtained at least on the basis of simple or specialist analysis;
- 4.0-4.3 – good, because data for at least half of the criteria were obtained by using at least a simple analysis (the lower limit of the range was assumed to be the weight of 0.8x5 libraries);
- 4.4-5.0 – very good, because the data for at least half of the criteria (min. 4) were obtained directly and for the rest of the criteria by simple analysis (the lower limit of the range was assumed to be the weight of 1x2 libraries plus the weight of 0.8x3 libraries).

To obtain a complete picture of the evaluation the following were calculated:

- The number of objects that had metadata for a given metadata element in a particular library – m_{le}
- the level of difficulty of obtaining data depending on weight features and number of resources in digital library

$$E_{1le} = \frac{w_{le} m_{le}}{m_l} \tag{3}$$

where:

m_l – number of all analyzed cartographic objects in a particular digital library.

Result

First, the method of obtaining metadata from individual digital libraries was analyzed, and individual features and weights were assigned to them (Table 3). In this way, the consistency of metadata in the analyzed digital libraries with the proposed set was checked (Tab. 2).
Table 4 presents the assignment of features and weights to individual libraries and metadata elements. It turns out that metadata typical of archival objects, such as: type of content, date, date range, main subject, rights, language, is easy to obtain (the level of difficulty of obtaining data (E_{1e}) is 3.0-5.0). Usually, they are collected in separate metadata fields, so they can be obtained directly. The only characteristic element for maps that can be obtained directly or with use of a simple analysis is the scale of the map. Other metadata characteristic for maps are not collected or it is difficult to obtain them (by specialist analysis or simple analysis) from academic libraries, while in the National Library of Poland they can be obtained by means of a simple analysis (geographic location) or specialist analysis (reference system, mapping methods, map format, orientation, source materials used to develop the map). Moreover, it was determined that it is easier to obtain metadata from the National Library of Poland. For metadata typical of all objects in the library, the results were very similar (5.3-6.0). For map-specific metadata in academic libraries, this level ranged from 1.0 at the Silesian University of Technology Digital Library to 2.9 at the Digital Library of University of Wroclaw, and 4.1 in the National Library of Poland. This value results from the fact that the more resources a given library has (Digital Library of University of Wroclaw - 1 733, and National Library of Poland 35 090), the better the metadata is described. There are strict guidelines on how to include detailed information in metadata.

Item	**Metadata element**	**Jagiellonian Digital Library**	**Digital Library of University of Wroclaw**	**Silesian University of Technology Digital Library**	**Maria Curie-Sklodowska University Digital Library**	**National Library of Poland**	E_{1e}
1	Type of content	1.0	1.0	1.0	1.0	1.0	**5.0**
2	Date	1.0	1.0	1.0	1.0	1.0	**5.0**
3	Date range	1.0	0.8	0.8	0.8	1.0	**4.4**
4	Main subject	0.5	0.5	0.5	0.5	1.0	**3.0**
5	Rights	1.0	1.0	1.0	1.0	1.0	**5.0**
6	Language	1.0	1.0	1.0	1.0	1.0	**5.0**
	E_{1l}	**5.5**	**5.3**	**5.3**	**5.3**	**6.0**	
		very good	**good**	**good**	**good**	**very good**	
7	Geographic location	0.0	0.0	0.0	0.0	0.8	**0.8**
8	Scale of map	1.0	0.8	1.0	0.8	0.8	**4.4**
9	Reference system	0.0	0.0	0.0	0.0	0.5	**0.5**
10	Mapping methods	0.0	0.8	0.0	0.8	0.5	**2.1**
11	Map format	0.0	0.5	0.0	0.8	0.5	**1.8**
12	Orientation	0.8	0.8	0.0	0.0	0.5	**2.1**
13	Source materials used to develop the map	0.8	0.0	0.0	0.0	0.5	**1.3**
	E_{2l}	**2.60**	**2.90**	**1.00**	**2.40**	**4.10**	
		bad	**bad**	**bad**	**bad**	**sufficient**	

Table 4: The level of difficulty of obtaining metadata for each library and for each element of metadata.

At the next stage, the level of difficulty of obtaining metadata was calculated depending on the number of objects that have metadata to check the completeness of the metadata. In this case, the results are shown for academic libraries and the National Library of Poland separately (Table 5.). As it was determined, academic libraries and the National Library of Poland still achieved similar results for the typical metadata of all objects in digital libraries (Academic Libraries - 5.2, National Library of Poland - 5.5), but the level of difficulty for map-specific metadata 0.7 for academic libraries and 3.0 for the National Library of Poland, which means the level of difficulty in obtaining metadata depending on the size of the resource for the National Library of Poland is over 4 times higher than for all academic libraries. It is also worth emphasizing that in the National Library of Poland 13 times more cartographic documents were analyzed than in academic libraries.

Item	Metadata element	Number of all objects that have data in		The level of dificulty of obtaining metadata depending on the number of objects	
		Academic libraries	National Library of Poland	Academic libraries	National Library of Poland
1	Type of content	2 638	35 090	**1.0**	1.0
2	Date	2 623	35 071	**1.0**	1.0
3	Date range	2 564	35 090	**0.9**	1.0
4	Main subject	2 536	18 285	**0.6**	0.5
5	Rights	2 131	35 090	**0.8**	1.0
6	Language	2 630	35 069	**1.0**	1.0
7	Geographic location	2 435	28 891	**0.1**	0.7
8	Scale of map	1 046	34 690	**0.3**	0.8
9	Reference system	0	32 271	**0.0**	0.5
10	Mapping methods	796	34 375	**0.1**	0.5
11	Map format	755	34 614	**0.1**	0.5
12	Orientation	110	34	**0.0**	0.0
13	Source materials used to develop the map	1	7 512	**0.0**	0.1
			E_{1le}	6.0	8.5

Table 5: The level of difficulty of obtaining metadata depending on the number of objects with metadata.

Discussion

This poor result on the level of obtaining map metadata from academic libraries is caused by several factors. First of all, map metadata is often supplemented by non-cartographers. In academic libraries, cartographic resources account for a small part of the total library resource, and a relatively small number of maps are collected and made available in digital form, and no one is focused on developing accurate and detailed guidelines for map description, i.e. where (in which specific metadata fields)) and how to describe cartographic studies. Therefore, every librarian can describe the same map in a completely different way. The situation is quite different in the National Library of Poland, which has employed cartographers, who meticulously supplement map metadata, for many years. Additionally, the National Library of Poland is an authority that is responsible for developing guidelines for developing metadata for all objects collected by libraries, including maps. In recent years, the following were developed: Instructions for cataloging cartographic documents (Szura, Krynicka 2012), and Cataloging regulations. Maps (National Library of Poland, 2018). This makes the metadata more complete and consistent.

It proved that metadata that are common to all objects in digital libraries, such as type of content, date, date range, main subject, rights, and language, are easy to enter by operators. This data is usually placed on maps and is easy to identify by librarians. The results for academic libraries and the National Library of Poland are similar.

It should be noted that the only information characteristic of maps and recognizable by catalogers in the academic libraries and the National Library of Poland is the scale. In the analyzed libraries it was supplemented for 99% of cartographic documents. In the National Library of Poland, geographical information is supplemented in the form of bounding box (geographical coordinates), whereas in other libraries location information is saved in form of names, which makes searching more difficult. Quite often, a given place has several forms, which hinders comprehensive search. This information can significantly affect the ability to find archival maps. Unfortunately, the National Library of Poland does not provide tools that would allow spatial search based on geographical coordinates. On the other hand, the Digital Library of University of Wroclaw, apart from the catalog, there is a map search engine based on coordinates (University of Wroclaw 2019).

Further research should concern the compliance of metadata values in individual libraries with the expected values i.e. verify whether what is entered by directories can be found by users. Such data could additionally help to determine not only whether we can find maps for our needs, but also to what extent metadata will help in answering the question how useful the maps we have found actually are.

Summary

In conclusion, university libraries have fewer map resources and describe these objects less appropriately. Quite often these objects are only a small part of the total resources they accumulate. The larger the collection, such as in the Digital Library of University of Wroclaw, the better the institution describes these data.

Based on the above research, it is worth looking for maps in the National Library of Poland, because its collections have the most complete metadata among all the analyzed libraries, cartographic documents are described in accordance with the developed guidelines, and it has the largest map collection. The National Library of Poland is a kind of authority in the description of metadata, among others maps and it collects more maps, so its map descriptions are better, more systematized, and complete.

References

Digital Library of University of Wroclaw (DLUW) (2019). In digital form, http://www.bibliotekacyfrowa.pl/dlibra

Digital Public Library of America (DPLA) (2019). In digital form, https://dp.la/

Dimitrov, B. (1998). Quality evaluation methods – a review. *Economic Quality Control* 13 (2): 117-128.

Europeana (2019). In digital form, http://www.europeana.eu/

International Organization for Standardization (ISO) (2014). 2014 Geographic information – metadata – Part 1: Fundamentals, ISO 19115–1. In digital form, www.iso.org/iso/home/store/catalogue_tc/catalogue_detailhtm?csnumber=53798

Jagiellonian Digital Library (JDL) (2019). In digital form, http://jbc.bj.uj.edu.pl/dlibra

Library of Congress – Network Development and MARC Standards Office (LoC MARC) (2019). MARC Standards. In digital form, http://www.loc.gov/marc/

Maria Curie-Skłodowska University Digital Library (MCSUDL) (2019). In digital form, http://dlibra.umcs.lublin.pl/dlibra

Moore, J. (2005). Digital Map Soup: what's Cooking in British Academic Libraries and are we helping our Users?. *LIBER Quarterly*, 15(1). DOI: http://doi.org/10.18352/lq.7800

National Library of Poland (2019a). API services. In digital form, https://data.bn.org.pl/

National Library of Poland (2019b). Catalogues of the National Library of Poland. In digital form, https://katalogi.bn.org.pl/discovery/search?vid=48OMNIS_NLOP:48OMNIS_NLOP&lang=en

National Library of Poland (2018). Przepisy katalogowania. Mapy [In English: Cataloging regulations. Maps]. In digital form, http://przepisy.bn.org.pl/przepisy-katalogowania/mapy#mapy

Park, J. (2009). Metadata quality in digital repositories: a survey of the current state of the art. *Cataloging & Classification Quarterly*. 47(3): 213–228.

Park, J., Tosaka, Y. (2010). Metadata quality control in digital repositories and collections: criteria, semantics, and mechanisms. *Cataloging & Classification Quarterly*. 48(8): 696–715.

Polona (2019). In digital form, https://polona.pl/

Polish Digital Libraries Federation (PDLF) (2019). In digital form, https://fbc.pionier.net.pl/

Silesian University of Technology Digital Library (SUTDL) (2019). In digital form, http://delibra.bg.polsl.pl/dlibra

Szura, R., Krynicka, M. (2012). Instrukcja katalogowania dokumentów kartograficznych [In English: Instructions for cataloging cartographic documents]. National Library of Poland. Warsaw, Poland.

The Dublin Core Metadata Initiative (DCMI) (2012). Dublin core metadata element set, Version 1.1. In digital form, http://dublincore.org/documents/dces

The Institute of Museum and Library Services Digital Collections and Content (IMLS DCC) (2019). In digital form, https://www.flickr.com/people/imlsdcc/

Tsou, M.H. (2002). An operational metadata framework for searching, indexing, and retrieving distributed geographic information services on the internet. In: Egenhofer, M. and Mark, D. (Eds),

Geographic Information Science (GIScience, 2002), *Lecture Notes in Computer Science*, 2478: 313-332.

University of Wroclaw. 2019. University of Wroclaw Electronic Catalog of Map Collections. In digital form: https://gaikk.bu.uni.wroc.pl/prod/mapa.html?lang=en

Youngblood, D. (2006). Map use across the disciplines. *Journal of Map and Geography Libraries*, *2*(2), 33-66.