

## Description schema of posters

***Since the first appearance of the electronic libraries more and more electronic collections have been built in the World Wide Web, where such documents come into prominence that do not have independency, not even in public libraries. Posters also belong to this group. In order that the digitized poster collections could be operated efficiently the use of metadata is essential because with their help the electronic documents can be successfully arranged and retrieved. So far metadata standard has not been constructed for the cataloguing of posters. In this research study the Poster Metadata Description Schema (PoMeDS) will be presented that we have developed. The metadata elements of (printed and digitized) posters can be described by this XML-based schema. As an XML-based schema, PoMeDS makes use of hierarchical container elements to group sub-elements or similar kinds of information belonging to the same area. It uses attributes to further refine or define the values contained by certain elements.***

Though books and journals are determinant document types in libraries since centuries, the collection interest of the libraries can include other document types too. Special collections can be organized from certain document types depending on the number of documents where small prints can be placed as well. Posters represent the most investigated and often borrowed part of this publication type.

The significance of the posters is that through them we can get to know more about social, political events and changes happening in industry, commerce and cultural life. An outstanding ephemera collection can be an indispensable source of historical and biographical research, particularly in the field of cultural, industrial and art history etc. For these prints were produced to communicate political fights of the centuries of our history, news and events of revolutions and counter revolutions and decrees to a mass of people. All these facts confirm that posters are components of library collections today and should remain components of the libraries in the future as well.

During conducting research in printed poster collections we can face difficulties, since it is not sure that it is enough for us to know the title and author of the work. In several cases it might be even necessary to have a look at the image itself on the poster. In order that a relevant work should be found we must look at more posters, which can be rather problematic in regard to storage and type of posters. On the one hand printed posters have a big size in general, therefore their viewing demands a lot of room. On the other hand their storage ideally takes place in chests of drawers, many times these documents, however, are put into letter files or are stored in boxes. In the latter case not only the viewing of posters is problematic, but also the orientation among boxes. Preparing the digitized version of posters and organizing them into an electronic collection can be a solution to this task.

It is definitely true for digitized poster collections that the use of metadata is essential for their operation as with the help of these data electronic documents can be efficiently organized and retrieved. The problem arises from the fact that libraries accepting the task of cataloguing apply a different set of metadata elements. Of course the most significant bibliographic data such as title, author, publication, number of pages are recorded, but the presentation of data which are typical of posters is undertaken only by a few libraries. Nevertheless the processing of library documents should be based on principles which are independent of country, language and library type. Concerning cataloguing great part of library documents the International Federation of Library Associations and Institutions (IFLA) ISBD program provides consistent principles, on the contrary the principles regulating the bibliographic description of posters are still missing.

Therefore I determined the data elements which are necessary for the bibliographic description and identification of printed posters, the order of data elements and the punctuation. Thus the principles of bibliographic description used for printed posters have been determined based on the ISBD rules. The main aim of the use of printed posters' secondary descriptive data is to determine such descriptive features regarding the poster document type that can help us to create a description meeting the requirements of ISBD „international standard bibliographic description“.

Bibliographic data of printed posters were completed by the metadata of digitized posters at the following step. Though several secondary descriptive data of printed and digitized posters are the same, the mode of recording digitized posters is different from that of recording printed posters, therefore new data elements come to the front such as system requirements needed for displaying an image, file characteristics of digitized image, technique used for creating the digitized image, etc.

Finally I analysed the storage of the determined data elements in computer-based systems. In order that digitized poster collections could be operated efficiently the use of metadata is essential because with their help the electronic documents can be successfully arranged and retrieved. To function properly in any computer-based system metadata should be encoded for machine readability and processing. Library catalog data are most often encoded in various MARC (Machine-Readable Cataloging) formats. Since appearance and form of electronic documents are different from those of printed documents, therefore MARC format can not be used appropriately for storing the required and sufficient descriptive data of electronic documents (but there were initiatives concerning this; for example, see MARC XML format). Storage of bibliographic data of electronic documents has been solved in several ways; for example, with the use of Extensible Markup Language (= XML). Both international and home practices reflect that XML schema is applied to storing metadata.

XML is an open source, nonproprietary encoding standard. It was developed in the 1990's as a more manageable subset of the much more vast SGML. Actually, it is a metalanguage employed for creating specific markup languages. XML has been widely adopted in publishing, e-commerce and other communities, including various cultural heritage metadata communities.

From the viewpoint of the library XML schema has an advantage, it also defines semantic rules which is typical of most cataloging standards, therefore several bibliographic rules can be formalized by XML schema. The order of data elements can be determined in schemas, which serves as basis for bibliographic description. In addition, we can define the restricted and the unrestricted data elements. Since the elements can be embedded into each other, therefore the closely related elements can be repeated. Another great advantage of the schema is that it does not support predefined element set, so metadata standards can develop their own schema in a flexible way. Actually, bibliographic descriptions can be made by XML schema which are exchangeable at international level, thus efficient search is also supported. Perhaps it might be the reason for the fact that XML schema is the most often applied standard for encoding metadata. Good examples of using XML are MODS (Metadata Object Description Schema), DC (Dublin Core) and VRA Core (Visual Resources Association Core).

Though there are descriptive schemas (such as DC, MODS, VRA Core), these do not contain every metadata necessary for the description of posters. These schemas can be completed by appropriate metadata necessarily, but I intended to work out a schema which can be used exclusively for the bibliographic description of posters and it can function as an effective search tool for posters and describe every special characteristics of posters. In this study the main characteristics of posters are described, those potential data elements are enumerated which can be used for their cataloguing. Finally a schema called PoMeDS (Poster Metadata Description Schema) is presented, which is working with the secondary descriptive data of printed posters and the metadata of digitized posters.

### **Bibliographic description of printed posters**

In conceptual definitions that can be found in lexicons, poster appears as a large, eye-catching, succinctly communicative printed material that is one-sided. However, from the public library point of view, poster belongs to ephemera. The category includes pamphlets, handbills, leaflets, broadsides, position papers, minutes of meetings, information sheets, bulletins, newsletters, posters, etc.

We can distinguish three types of posters based on their formal features: namely textual, picture and picture with text ones. In the following, I classify the latter two types in one group, which are the graphic posters.

In the case of textual posters written content has an importance, while regarding graphic posters the emphasis is put mainly on the represented picture, although the text complementing the picture can also have a significant role.

Whether we talk about textual or graphic poster, it always has a purpose: attracting attention, providing information or persuasion by means of picture and/or text. Past events are taking shape mosaic like from the pictures and the text of the posters, therefore through them we can get to know more about the social, political events and the changes happening in industry, commerce and cultural life. In addition to this, graphic posters represent the precious sources of artistic life as well. As a result of this, posters are at the same time documents of cultural history and history of art forming a part of visual culture and they also reflect a historical period by virtue of their content.

As graphic posters emerged, they immediately became part of the public collection. This is due to the librarian perception that states that "it is not the task of the present to decide what will be important for the future, therefore everything should be preserved that gets into libraries." However, libraries' task goes beyond the gathering of documents, because in order to be able to use a collection, documents have to be catalogued and made retrievable.

There are not specific rules of procedure and standards yet for the description of posters appearing in printing. The ISBD/NBM (International Standard Bibliographic Description for Non-Book Materials) or ISBD/G (General International Standard Bibliographic Description) can be one of the bases for standard bibliographic description of printed posters. We can find more guides to catalogue graphic materials or pictures. These guides, standards regulate the bibliographic description of many document types at the same time, thus neither are only created to catalogue posters.

Owing to the use of various guides and rules, libraries have not catalogued posters in a consistent way. The most important bibliographic data like: title and statement of responsibility, publication and number of pages are recorded everywhere, but there might be differences in respect of special aspects expressing the characteristics of posters. Taking into consideration that nowadays besides the traditional library materials, these printed materials come into more and more prominence, therefore it was necessary to create a metadata element set that is only suitable for the bibliographic description of posters. Accordingly, one of my research aims was to determine the data elements which are necessary for the bibliographic description of posters, to make it possible to create ISBD-based description thanks to these. ISBDs are based on internationally accepted principles. The primary aim of ISBDs is to support the international use of bibliographic information by making records from different sources interchangeable. So the bibliographic records produced in one country can be easily accepted in library catalogues or other bibliographic lists in any other country. The another aim of ISBDs is to help the international interpretation of records and the reduction of language barriers, so records produced for users in one language should be interpreted in other languages as well.

When elaborating the bibliographic description of graphic posters, detailed standard was missing, hence I relied on guidance and other institution's practice. As my goal was to make standardized description about posters based on the data elements I defined, therefore the ISBD (NBM): International Standard Bibliographic Description for Non-Book Materials standard, MSZ 3424/1. Bibliographic description. Books standard and a guideline with title „Bibliographic description of stills“ served as basis for determining and defining the appropriate data elements. Apart from all these I have taken into consideration the guidelines of the publication entitled "Graphic materials: rules for describing: original items and historical collections" used by the

Library of Congress. In addition to standards and guidelines, the practice of the Library of Congress (LC), National Széchenyi Library and the University of Debrecen University and National Library provided a good basis for determining the proper data elements which can be used for the description of posters. Studying the standards, guidelines and practices, necessary data elements for the description and identification of printed graphic posters were determined, as well as the order of the data elements and the punctuation.

In the determination of posters' secondary data, primarily, I have used ISBDs that are based on internationally accepted standards, therefore general rules were determined such as:

- structure of the description of printed poster, punctuation
- sources of information used for description
- language and script of the description
- capitalization
- misprints.

Twenty-seven data elements are defined in seven areas:

- Title and statement of responsibility area (title proper, general material designation, parallel title, other title information, statement of responsibility)
- Edition area (edition statement, statements of responsibility concerning the edition)
- Publication, distribution area (place of publication, name of publisher, date of publication, place of manufacture, name of manufacture)
- Physical description area (specific material designation and extent, other physical details, dimension, accompanying material statement)
- Series area (title proper of series, parallel title of series, other title information of series, statements of responsibility concerning the series, number within the series, title proper of sub-series, parallel title of sub-series, other title information of sub-series, statements of responsibility concerning the sub-series, number within the sub-series)
- Note area
- Distribution data (manufacture number)

## Process of digitized poster

For a long time, only paper-based posters existed. Thanks to libraries' digitization projects that included a growing number of document types, however, digitized versions of these documents also emerged. Digitized posters slowly became part of electronic collections as well.

In consequence of this, electronic documents and digitized posters should be made retrievable by using the suitable data. However, these documents have unique characteristics, which no longer can be described by the traditional library categories. Metadata provide solution to these problems. Therefore bibliographic data of printed posters were completed by the metadata of digitized posters. Though several secondary descriptive data of printed and digitized posters are the same, the mode of recording digitized posters is different from that of recording printed posters, therefore new data elements come to the front such as system requirements needed for displaying an image, file characteristics of digitized image, technique used for creating the digitized image, etc. Taking all these into consideration, the maximum of descriptive data elements has been determined regarding both printed and digitized posters. Arising from this, there are essential data elements for the identification of documents and there are other data elements which can be omitted from the bibliographic description.

To function properly in any computer-based system metadata should be encoded for machine readability and processing. The process of encoding library catalog data started in 1960's. In 1962 the Library of Congress launched its MARC project. As a result of this initiative the first MARC format, the LCMARC appeared. MARC is used for communicating and sharing bibliographic data among library systems and bibliographic utilities. Different countries have developed their own implementations of MARC format adjusted to their national cataloguing practice. [18] For example USMARC, Canadian MARC and HUNMARC etc. were created in this way. Later USMARC (i.e. the national MARC format of the Library of Congress) and Canadian MARC (i.e. the national MARC format of the National Library of Canada) merged to become MARC21. Today there are continuing movements for the harmonization of MARC format among many countries. Advancement in storing metadata took place when the World Wide Web and several electronic documents appeared and spread. Today, most of the major cultural heritage metadata schemas have been expressed as XML schemas.

## Poster metadata description schema

Although there are descriptive schemes (such as DC, MODS, VRA Core), these do not contain every metadata necessary for the description of posters. These schemes can be completed with appropriate metadata necessarily, but I wanted to work out a schema which can be used exclusively for the bibliographic description of posters and can be used as an effective search tool for posters and every special characteristic of posters can be described by this schema. Thus I have developed the Poster Metadata Description Scheme (PoMeDS). It is an XML-based scheme storing the metadata of the posters and working with such a set of metadata elements.

### Metadata

Elements applied in the schema can be divided into two major groups by their use in the bibliographic description of printed posters or digitized posters. There are metadata in PoMeDS which appear in both groups; for example, the <titleProper> and

the <subTitle> elements. Primarily because the title proper and the subtitle of the digitized document often correspond with those of the printed poster.

Metadata of printed and digitized posters can be categorized on the basis of the functions they fulfill. Majority of metadata applied in the schema are descriptive metadata. These metadata can be used in the bibliographic description and retrieval of posters, namely the poster will be identified by them. Descriptive metadata are; for example, metadata used in the bibliographic description of printed posters or elements of heading. Metadata of digitized posters are more complicated, besides descriptive metadata we can find other types of metadata. Metadata can be grouped by Steve J. Miller in the following way:

- descriptive metadata: title (titleProper, generalMaterialDesign, subTitle); satOfResp (person, corporate); date; fileCharacteristics (fileName, identifier, format, fileSize); relations (relation, note); origDocBiblDescr (bibId, opacUrl); cataloger (person)
- administrative metadata:
  - Technical and preservation metadata: fileCharacteristics (fileName, identifier, format, fileSize, compressionRatio, location); technique (modelName, modelNumber, modelsSerialNo); polarity (polarityType); generation (generationType);
  - right metadata: right (rightHolder, note);
  - use metadata: systemRequirements (hardwareRequirements, softwareRequirements, peripheralsRequirements); access (accessRestriction, accessCondition).

### ***Elements and attributes***

In PoMeDS the meaning of each metadata can be specified in several ways. Exploiting the hierarchical structure of the XML schema, additional elements were defined within metadata to be specified which have already determined the meaning of the given metadata more concretely. For example <accessRestriction>, <accessCondition> elements appear also within <access> metadata. It is obvious that with the help of each additional element we can define different data. At the same time the PoMeDS uses XML attributes also for specifying each element, for designating certain codings or controlled vocabularies and for other similar purposes. These attributes occur in start tag. Elements can include more than one attribute too. In the case when the values of elements must be specified, the use of attribute is a better solution than the declaration of qualifiers or new elements.

PoMeDS is a hierarchical schema which often uses container elements. Such container element can be <posterCollection> element which contains the <originalPoster> and the <digitizedPoster> sub-elements. The <edition> element is container element too because the <editionStatement>, the <pointSpaceDash> and the <statOfRespRelatingToTheEdition> sub-elements are put inside the <edition> element.

This schema uses attribute groups which include set of several attribute declarations. An attribute group can be defined once as a child of the schema elements and can be referred to within any element by using its declared name. Such attribute group is called PoMeDS involving date, hypref, source, modify, language attributes. The second one is authority attribute group which includes the authoritydate, ID, authorityhypref and authority source attribute declarations. Almost every element of the schema includes PoMeDS attribute group. However, the authority attribute group can be used in the case of elements of heading.

At the bottom level there are defined elements having PoMeDSstring, PoMeDSinteger, PoMeDSanyURI or PoMeDSgYear types. It means that an element has been created with the extension of the definition of the string, integer, anyURI or gYear simple types. Thus the values of earlier mentioned types are allowed as the content of elements. Optionally date, source, hypref, modify, language attributes can be given which belong to set of PoMeDS attribute group.

As the order of elements is defined, the PoMeDS gives rules about the order of data elements. In XML schema the order of elements can be defined by using the „sequence“ compositor. At the same time one element can be chosen out of more elements with the help of „choice“ compositor. In the schema the occurrence of a given element can be defined (e. g. it can be mandatory or optional etc.). In addition to, different elements can be embedded into each other. All these features result in substantial flexibility inside the schema. Owing to these solutions, issues will be solved easily such as problem of parallel data or bibliographic description of one or more, same or several places of publication inside a publication, distribution area.

### ***Punctuation marks***

All bibliographic description standards determine rules about the order of the data elements and the punctuation marks which precede data elements. This provides the recognizability of data elements inside an area and the international understanding and interpretation of the bibliographic description. Therefore PoMeDS gives rules about the order of the punctuation marks which precede the data elements. In the schema only those punctuation marks (i.e. point-space-dash, new indent) were defined which separate the areas and thanks to these marks (point, semicolon, equals sign, diagonal slash, plus sign, colon, comma, round bracket) the data elements can be recognized. These marks have been defined as an independent, empty XML element in the schema, so a complexType element was used as a type definition in the XML schema which does not declare the content model.

Concerning punctuation marks it is the cataloger's task to determine those marks which are associated with concrete bibliographic data in the description, such as; for example, the three dots which indicate that text has been omitted in some areas.

### ***The structure of the PoMeDS***

Supporting the bibliographic description of digitized posters, the PoMeDS is composed of two top level elements (<originalPoster> and <digitizedDocument>): we can process the original document with the help of one element and the another

element complements the previous one with the data typical of the digitized version. The <originalPoster> element includes metadata of printed posters. These data are complemented with metadata characterizing the digitized version which can be found in the digitizedDocument element.

The description of the original document (originalposter) is divided into twelve major units according to the areas and the headings of the bibliographic description:

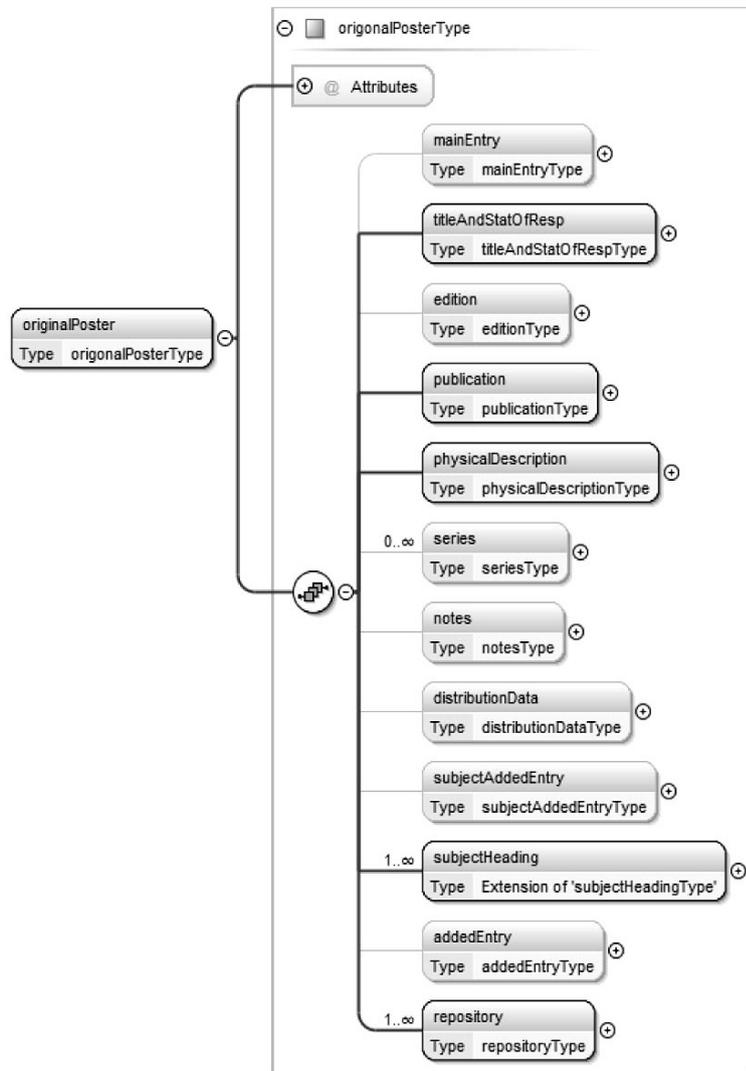


Fig 1 Diagram of originalPoster

Concerning digitized posters, data consist of two main parts: originalPoster, digitizedDocument.

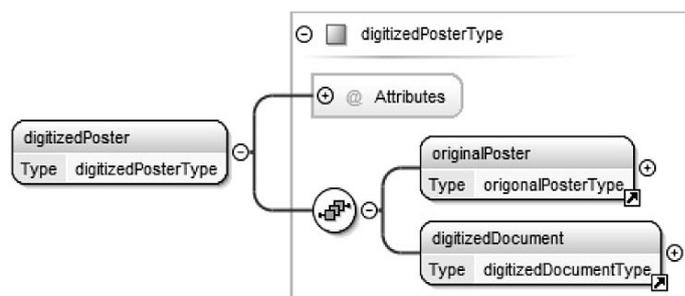


Fig 2 Diagram of digitizedPoster

The originalPoster is made up of elements discussed before. These data are complemented by metadata which characterize the digitized version and they can be found in the digitizedDocument element:

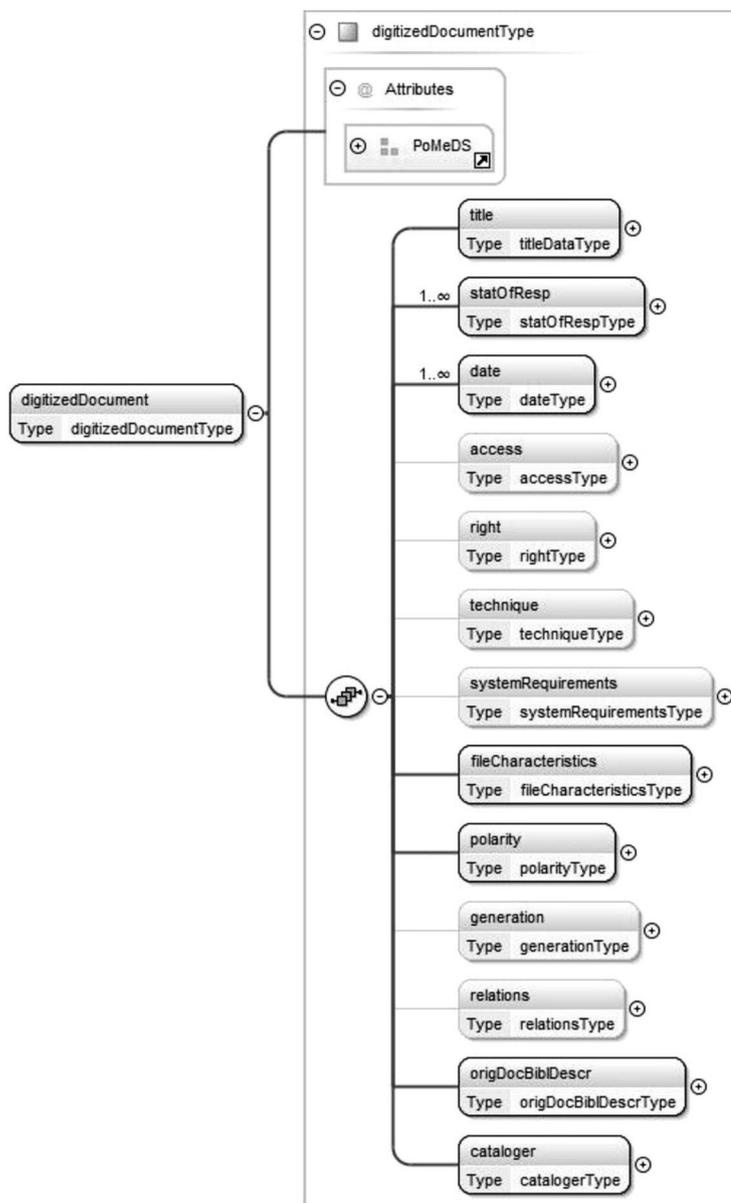


Fig 3 Diagram of digitizedDocument

In this study the Poster Metadata Description Schema will be presented that we have developed. The PoMeDS works with the secondary descriptive data of printed posters used in standard ISBD descriptions and the metadata of digitized posters determined by us. With the help of the schema, we are able to catalogue printed posters, complete the description with appropriate headings, as well as give the metadata of the posters' digitized version. Consequently every special characteristics of posters can be described by this schema. Thanks to this, the spread of PoMeDS schema would support to access printed and digitized posters in a better way. However, to use the schema effectively in practice, it should be interoperable, harvestable and shareable. Making these conditions possible would provide the basis of further research. Furthermore web search engines can interpret and work seamlessly with XML format, therefore the elaborated schema can contribute to further research which provides opportunity for developing "smart" or "intelligent" library search engines by exploiting artificial intelligence research.

**References**

E. W. Betz: *Grafic materials: rules for describing: original items and historical collections*. Washington D.C.: Library of Congress, 1982.

# BIBLIOGRAFIA/INFORMAČNÉ VZDELÁVANIE

## BIBLIOGRAPHIA/INFORMACIONE ASZTUDYALÁS

- Gy. Pogány: "Különgyűjtemények". Könyvtárosok kézikönyve 3., T. Horváth, I. Papp, eds. Budapest: Osiris, 2001, pp. 179-207.
- I. Varga: Állóképek bibliográfiai leírása: útmutató. Budapest: OSZK-KMK, 1981.
- ISBD: International standard bibliographic description. Standing Committee of the IFLA Cataloguing Section, Consolidated ed. Munich: Saur, 2011.
- ISBD(G): General International Standard Bibliographic Description. London: IFLA Int. Office for UBC, 1977.
- ISBD(NBM): International Standard Bibliographic Description for Non-Book Materials. London: IFLA Universal Bibliographic Control and International MARC Programme British Library Bibliographic Services, 1987.
- J. Tóvári, B. Szabó: Metaadat tárolási technológiák. Eger: Eszterházy Károly Főiskola, 2011.
- M. Cseh: "Az utca művészete: képes plakátok Benczúr Gyulától Konecsni Györgyig". A nemzeti könyvtár kisnyomtatványtára, M. Cseh Eds. Budapest: Osiris, 2002, pp. 12-35.
- M. El-Sherbini, G. Klim: "Metadata and cataloging practices". The Electronic Library, vol. 22, no. 3, pp. 238-248, 2004.
- M. J. Young: XML lépésről lépésre. Bicske: Szak, 2002.
- M. Németh: "Aprónyomtatványok az Országos Széchényi Könyvtárban 1918-1944". Az Országos Széchényi Könyvtár évkönyve, 1984/1985, pp. 205-234, 1985.
- M. Némethi-Takács: Plakátok leíró adatai és azok tárolása XML sémában. Debrecen: Debreceni Egyetem, 2017.
- MSZ 3424/1. Bibliográfiai leírás. Könyvek
- S. J. Miller: Metadata for digital collections. New York: Neal-Schuman, cop. 2011.

**Margit Némethi-Takács**

(Department of Library Informatics Faculty of Informatics, University of Debrecen) ■