Metadata for Integrated Catalogue: Presidential Library experience

- Olga N. Zhlobinskaya[1]
- Yulia G. Selivanova [2]

Corresponding Author: Yulia G. Selivanova

Abstract

Integrated catalogue of the Presidential library is described, including heterogeneous resources (library, archival, museum) and collection-level descriptions for digital collections. Basic principles of the integrated catalogue are given such as: common metadata standard for various types of resources and their aggregations; consistent methodology for describing and indexing, taking into account the characteristics of each type and type of resources; use of common authority files to ensure the control of access points for names and subjects; and common search interface that provides access and viewing of diverse resources, digital collection being considered as one of types of resources. Main focus is using national metadata format RUSMARC as common metadata structure standard for various types of resources as well as for collection-level descriptions. It is shown that RUSMARC not only provides precise description and differentiation of various types of resources and their aggregations but also ensures structural compatibility of the descriptions and enhanced functionality of user search interface. RUSMARC metadata provides implementing single search box for all types of resources in the catalogue, using specific characteristics of each type of resource to further refine the search results.

Keywords: *Presidential Library; Metadata structure; Authority file; Integrated catalogue; Library catalogue; Digital collections; Archival materials: RUSMARC*.

Introduction

The Presidential Library in Saint-Petersburg, one of the three national libraries in Russia, was established in 2009, the statutory goal of the Library being the creation of a national repository of digital information resources (both digital born and digital copies) on the history, theory and practice

1   **Olga N. Zhlobinskaya** *is Senior specialist of the Library Cooperation Department at the Presidential Library, Saint-Petersburg, Russia.*

2   **Yulia G. Selivanova***, PhD is Director of the Library Cooperation Department at the Presidential Library, Saint-Petersburg, Russia. Email: selivanova@prlib.ru*

of Russian statehood, and on the Russian language as the state language of the Russian Federation.

Original library's collection was based on digitized materials from Russian State Historical Archive. Currently main sources of acquisition are state and regional archives (including Russian State Historical Archive, State Archive of Russian Federation, Leningrad Regional State Archive in Vyborg, etc.), national libraries and major research libraries of the country. The Presidential library work in cooperation with regional libraries and research institutions in Russia and abroad, Russian Geographical society, other non-profit organizations, private collections.

By 2024 the collection of the Presidential Library includes about 1 100 000 000 items; it is an integrated corpus of digital resources including text, visual resources, photos, maps, audio- and video recordings, multimedia resources, virtual tours, etc. More than a half of the collection are archival materials, and also the collection includes a number of museum objects (mainly their visual representations).

The library collection development follows its statuary goal and is based on "collection-oriented principle". At the highest level the collection of the Presidential Library is organized into four basic thematic blocks which started as digital collections: "State authority", "Territory of Russia", "Russian people", and "Russian language", revealing various aspects of the concept of Russian statehood. In addition, the Library is creating thematic digital collections dedicated to certain periods and events of Russian history, regions of the country, the activities of famous public figures and politicians.

Every resource in the library's collection is included into one or more collections. Thus, in general, the collection of the Presidential Library is a complex of digital collections, which are considered not only as means of organizing and managing resources but also as additional instrument of navigation in the collection as a whole.

All the resources are described with RUSMARC format (national implementation of international format UNIMARC). But the above features of the collection of the Presidential Library, namely great number of non-library resources (i.e. archival materials and images of museum objects) and "collection-oriented" principle of the collection development, required special solutions on metadata strategy for the library catalogue.

By 2009 when the President Library was established none of the libraries in Russia had any experience of developing integrated catalogue including various types of materials (namely library and archival materials) as well as developing great number of digital collections and implementing full-scale search and navigation across the collections. As for international practice, catalogues including library and archival materials have been developed since 1970-s in the Library of Congress (USA) and in the Library and Archives Canada, but they were based on MARC 21 rather than on UNIMARC. So, there was an urgent need to investigate possible approaches to the above tasks with minimal labor intensity, financial and intellectual costs.

Literature Review

Archival and library materials are quite similar in that they document human history but

normally they are not listed in library catalogues. Among the reasons to be mentioned is different arrangement of materials as well as highly hierarchical structure of the archival description. Unlike bibliographic record, which usually focuses on an individual manifestation of a published work, an archival description is highly hierarchical and normally involves a complex group of interrelated unique materials whose shared provenance and hierarchical arrangement must be shown very clearly.

Major international content and encoding standards regulating archival description and its machine-readable presentation at that time were: General International Standard Archival Description (ISAD (G) (International Council on Archives, 2000); ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families (International Council on Archives, 2004); Describing archives: a content standard (DACS) (Society Of American Archivists, 2019), and Encoded Archival Description (EAD).

ISAD (G) provides a set of general provisions to ensure the creation of consistent and self-explanatory archival descriptions and facilitate the retrieval and exchange archival information. The standard formulates the basic principle of archival description – archival description proceeds from the general to the specific. Also, hierarchical model of the levels of arrangement for the fonds and its constituent parts (fonds, series, file and item with possible sublevels) is presented. The elements of the description are arranged into seven areas of descriptive information.

ISAAR (CPF) defines data elements and general provisions for the creation of authority records for names of creators (corporate bodies, persons or families). ISAAR (CPF) also takes into account the experience of libraries in authority control.

DACS provides a set of rules for describing archival materials, personal papers, and manuscript collections. These rules allow for the description of archival materials at any level. The documentation includes rules for the creation of authority records.

EAD is an international standard for the presentation of archival descriptions in machine-readable form. It provides a standardized electronic description of archival and handwritten collections and is widely used for online access to archival collections around the world.

On the national level in Russia, we did not have (and still do not have) national standard for archives, just recommendations and general regulations issued by the Federal Archival Agency, and practice used in the country might vary from one archive to another.

In the library community, major standards used are ISBD (in Russia – Russian Cataloguing Rules and national standard for bibliographic description GOST R 7.0.100–2018) and MARC formats. Both ISBD and MARC historically were not supposed to be used for archives and in general they do not allow to create detailed archival description. Since the 1970s, the Library of Congress has been trying to expand the scope of the MARC format for describing archival materials. MARC AMC (MARC for Archive and Manuscript Control) has been developed which was the first attempt to integrate the presentation of library and archival materials in a single machine-readable format. However, this attempt was not successful with the archival community, since the format was focused more on the rules for describing published documents and did not provide a

hierarchical archival description.

Later, the MARC21 format for describing archival documents was used in bibliographic institutions if there were book and archival documents in the collections. The Library of Congress includes a bibliographic record in MARC21 format for the fonds as a whole, while a hierarchical archival description is generated using EAD.

In general, MARC record gives just a horizontal view of the object so it not the best instrument to describe archives. This problem is partly solved in UNIMARC and UNIMARC-based formats (including RUSMARC) which have the mechanism of building comprehensive multilevel descriptions and establishing links between different levels but UNIMARC did not provide archival-specific data elements.

Methodology

In 2010 Presidential Library established Working group to develop the principles and approaches to integrated presentation and access to library, archival, museum resources in the catalogue in compliance with modern international standards.

The scope of the Working group was: to study national and worldwide practice of archival description as well as international standards and machine-readable formats in the field; to determine specific requirements for archival description in the library catalogue; to develop the most effective way of treatment of archive materials; to determine data elements for archival materials of resource missing in RUSMARC; to make appropriate updates to RUSMARC format.

Based on the analysis of national and worldwide practice of archival description as well as international standards and machine-readable formats in the field (including ISAD (G), ISAAR (CPF), and EAD tag library, practice of Library of Congress and the Library and Archive Canada) the Working group made recommendations to create an integrated catalogue covering various types of resources which would enable the users to locate resources of interest more easily. They also agreed that RUSMARC, with some improvements, could be used for archival descriptions.

Comparative study of RUSMARC and ISAD (G) showed that the format lacked some data elements which were essential for archives. In 2012 some changes and additions in RUSMARC were made which we supposed to be necessary to present archival materials. After These updates included: specific value in the record label to specify archival control for the materials which allowed to implement different options of search, facet navigation and displaying search results to users; additional data elements for treatment of dates in archival description as well as explicit indication of the level of the description.

Results and Discussion

The Working group outcomes and RUSMARC updates were essentially the basis for the integrated catalogue of the Presidential Library. Description of archival materials for the integrated catalogue in the Presidential Library is based on the following principles:

1. Archival materials are coded in the record label which allows to implement specific options

for search, facet navigation and displaying search results. In defining archival materials, the most important is not the type of resource (text, photo, etc.) but the way by which materials are arranged, physically organized and handled by the custodian.

2. Multilevel archival description may be created with any desired level of detail – it may be collection level or a set of records for all desired levels (fonds – series – file – item – component part). Each level contains only information relevant to that specific level.

3. Machine-readable record for archival materials contains data elements specific for archival documents and additional data elements defined in RUSMARC, e.g. coded data or material-specific fields for photographs, sound records, maps, etc. may be used where appropriate.

4. All controlled access points (personal names, corporate names, family names, place names, subject headings) at any level of description are controlled with the authority files that are common for library and archival materials.

The above principles and practice of the Presidential Library form the basis of UNIMARC Guidelines for Archival Materials (Zhlobinskaya O., 2023*)*. The Guidelines document the provisions for using UNIMARC in presenting archival materials and give examples of full UNIMARC records for archives. Also, Table of correspondence of EAD elements/attributes and UNIMARC fields/subfields is provided.

Workflow of describing archives in the Presidential library proceeds from the general to the specific, which corresponds to basic theoretical principle of archival description, i.e. the *"principle of respect des fonds"*. The process begins with creation of collection level record for the fonds. Then linked records are created for series, file(s), item(s) – depending of the level of details desired, linked to the above level. At every level the description is indexed with Library Bibliographic Classification and subject headings. The workflow is presented in the edition "Best practices for describing and indexing of digital copies" (Завьялова Л.В. et al., 2014) published by the Presidential library.

Using RUSMARC for describing archival materials proved to be very effective. Hierarchical links provided in RUSMARC help to keep hierarchical structure of archival description. On the other hand, extensive set of coded data and note fields and controlled access points provided in RUSMARC help to enrich archival description to make it more precise.

Archival and traditional library materials complement each other, and their co-existence in the library catalogue makes it more efficient and user-friendly; it also helped to enhance users' search interface. After the search query is processed the user may choose specific category of materials to be viewed – library or archival. Facets used for navigation to refine the search results vary depending on the type of materials. In particular, for archival materials there is an option to browse records at specific archival level – fonds, series, file or item, or to switch to another level when necessary.

In 2024 the Presidential library plan to initiate a project to evaluate effectiveness of the integrated catalog in improving user search experiences. It is planned that the methodology will include user surveys as well as some quantitative analysis details of which have yet to be determined.

Another characteristic of the collection, which determines the features of the catalogue is "collection-oriented principle" of developing the library's collection.

Digital collections act as structural units of the Library collection and are primarily considered as means of organizing and managing resources. But, in addition to performing an administrative function, collections can also act as one of the information search instruments as they give a general idea of the structure of the library's collection as a whole and the repertoire of resources in the library; provide top-level navigation in a heterogeneous collection of the library; help to provide information about the general provisions for access and usage of the resources of the collection.

In the Presidential library a description of the digital collection as an independent information object is created and included into the library catalogue, with links established to descriptions of individual resources in the collection. Such collection-level descriptions reveal information about the existence of a digital collection as a complex set of information resources, its characteristics and the availability of the resources included in the collection. According to The Framework of Guidance for Building Good Digital Collections, "collections should be described so that a user can discover characteristics of the collection, including scope, format, restrictions on access, ownership, and any information significant for determining the collection's authenticity, integrity, and interpretation." (NISO, 2007) Descriptive metadata of the collection in the Presidential library catalogue include following information: general topic of the collection; sources of acquisition of the resources; types of the resources in the collection (text, audio, video, etc.); number of resources in the collection; scope and extent of the collection; languages used in the collection resources; annotation; organizations that participated in the collection development (e.g. organizations which provided materials for the collection); the copyright status of collection resources; restrictions on access and usage of the resources; structure of the collection, and whether the described collection is a subcollection of another collection.

Also, it is very important, that the collection-level description documents explicitly the relationships established for the collection. The links are established with normal instruments provided in RUSMARC, these can be horizontal (associative) relationships, based on common topics, creators, etc., or vertical (hierarchical) – between collections and subcollections of different levels, or between collections and individual resources, both within the digital library and outside it. These relationships can be used as an additional means of navigation in the electronic library collection. All the collection-level records are also indexed with subject headings and Library Bibliographic Classification, and common authority files are used to control access points for all types of resources in the catalogue, including collection level. Controlled access points include personal and corporate names, place names, titles of works, subjects, and genre headings.

All the collection metadata is presented also in RUSMARC format. Thanks to the structured and standardized format of the machine-readable record, descriptions at the collection level enable extended functionality of user search interface – they are used for navigation and browsing arrangements, for preliminary selection of collections based on a general description, for grouping the record retrieved, for preliminary assessment of the relevance of the search and refinement search results, to search within the collection. Collection metadata has specific weightings applied when calculating search relevance, specific facets are used for navigation, etc.

Conclusion

Integrated catalogue of the Presidential library includes descriptions of library, archival, museums resources, as well as collection-level descriptions for digital collections as higher-level information resource which are used as independent search objects.

Basic principles of the integrated catalogue are: common metadata standard for various types of resources and their aggregations; consistent methodology for describing and indexing, taking into account the characteristics of each type and type of resources; use of common authority files to ensure the control of access points for names and subjects; and common search interface that provides access and viewing of diverse resources.

RUSMARC is used as common metadata structure standard for various types of resources as well as for collection-level description. Using common metadata standard ensures structural compatibility of the description, while providing specific description and differentiation of the description of various types of resources and their aggregations to ensure enhanced functionality of user search interface.

References:

Encoded Archival Description Tag Library- Version 2002 (EAD Official Site, Library of Congress). (n.d.). https://www.loc.gov/ead/tglib.

International Council on Archives (2000). ISAD (G): General International Standard Archival Description. https://www.ica.org/app/uploads/2024/01/CBPS_2000_Guidelines_ISADG_Second-edition_EN.pdf.

International Council on Archives (2004). ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families. https://www.ica.org/app/uploads/2023/12/CBPS_Guidelines_ISAAR_Second-edition_EN.pdf.

NISO (2007). A Framework of Guidance for Building Good Digital Collections. https://www.niso.org/sites/default/files/2017-08/framework3.pdf.

Society Of American Archivists. (2019). Describing archives: A content standard. https://files.archivists.org/pubs/DACS_2019.0.3_Version.pdf.

Zhlobinskaya, O. (2023). UNIMARC Guidelines for Archival Materials. https://repository.ifla.org/bitstream/123456789/3122/1/Guidelines_10_Archival_Materials.pdf.

Zavyalova, L. V., Maskhulia, T. L., Selivanova, Yu. G., & Stegaeva, M. V. (2014). Methodological recommendations for cataloging and indexing electronic copies of documents (E. D. Zhabko, Ed.). Presidential Library. https://www.prlib.ru/item/343460.