



ISSN: 2350-0328

## International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 6 , June 2020

# Format for Bibliographic Information Exchange

Jasurbek Atadjanov

Head of department, Department of Billing System Development, Uztelecom Stock Company, Tashkent, Uzbekistan

**ABSTRACT:** In this paper, it was described format of exchanging bibliographic information and its full text between information systems. This format consists of two main parts: header (bibliographic data) and body (context of full text). It is based on SOAP (Simple Object Access Protocol is a messaging protocol specification for exchanging structured information in the implementation of web services in computer networks) protocol. Bibliographic data can be any MARC (UNIMARC, MARC21, RUSMARC, etc.) formats and also full text can be any digital file formats (MsWord, excel, pdf, HTML, audio-video, etc). The main difference of this format from ISO-2709 (ISO 2709 is an ISO standard for bibliographic descriptions, titled Information and documentation – Format for information exchange) is related to supporting export-import of the full text of bibliographic data.

Expressing of bibliographic data and its full-text information in this format we have to do several steps:

- convert bibliographic data into XML (Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.);
- reading full-text content and by base64 algorithm converting into simple text;

This format can be used not only between library information systems but also in other information systems to exchange bibliographic data.

**KEYWORDS:** MARC, bibliographic data, SOAP, XML, Base64

### I. INTRODUCTION

There are many standards and formats to exchange bibliographic data (BD). Many of them can not support to exchange an bibliographic data with attachment (full text of this BD). Today there are many electronic libraries and virtual libraries which users can remote use their electronic catalog. Main part of these electronic catalogs are consists of electronic resources. When we exchange that typed BD it is important to include their full text. Full text may be as text format (TXT, PDF, DOC and etc.), as mixed media formats.

MARC is an acronym, used in the field of library science, that stands for Machine-Readable Cataloging. The **MARC standards** consist of the MARC formats, which are standards for the representation and communication of bibliographic and related information in machine-readable form, and related documentation. The record structure of MARC is an implementation of ISO 2709, also known as ANSI/NISO Z39.2. MARC records are composed of three elements: the record structure, the content designation, and the data content of the record [1]. Every MARC formats consists of fields (which numbered from 001 to 999), sub fields (0..9, A..Z) and data of this field. MARC fields maybe simple fields and multiple fields.

SOAP, originally defined as **Simple Object Access Protocol**, is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks. It relies on eXtensible Markup Language (XML) as its message format, and usually relies on other Application Layer protocols (most notably Remote Procedure Call (RPC) and HTTP) for message negotiation and transmission. SOAP can form the foundation layer of a web services protocol stack, providing a basic messaging framework upon which web services can be built. This XML based protocol consists of three parts: an envelope - which defines what is in the message and how to process it - a set of encoding rules for expressing instances of application-defined datatypes, and a convention for representing procedure calls and responses [2]. SOAP with Attachments (SwA) or MIME for Web Services refers to the method of using Web Services to send and receive files using a combination of SOAP and MIME, primarily over HTTP.

The SOAP architecture consists of several layers of specifications for message format, message exchange patterns (MEP), underlying transport protocol bindings, message processing models, and protocol extensibility.



ISSN: 2350-0328

## International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 6 , June 2020

### II. SCOPE OF RESEARCH

In this article I tried to describe how to convert BD with Attachments into SOAP. This algorithm maybe use to exchange BD between automated library systems.

### III. METHODOLOGY

In the below this article we have described MARC formats and SOAP protocol specification. Lets see example a BD which is described in MARC21 format.

245^A Spring in Action  
100^A Craig W.  
700^A Ryan B.  
020^A 1-932394-35-4  
260^C 2005

In the SOAP form of this BD we will use the MARC21 fields as SOAP tags, the sub fields as sub tags, data as value of this sub tags. So, BD will be the following form, when we convert it into SOAP.

```
<?xml version="1.0"?>
<soap:Envelope xmlns:soap="http://www.w3.org/2001/12/soap-envelope"
soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding">

<soap:Body xmlns:m="http://www.natlib.uz/carmat">
<m:mf245>
<m:msA>Spring in Action</m:msA>
</m:mf245>
<m:mf100>
<m:msA>Craig W.</m:msA>
</m:mf100>
<m:mf700>
<m:msA>Ryan B.</m:msA>
</m:mf700>
<m:mf020>
<m:msA>1-932394-35-4</m:msA>
</m:mf020>
<m:mf260>
<m:msC>2005</m:msC>
</m:mf260>
</soap:Body>
</soap:Envelope>
```

In this example I have added before MARC21 field number “mf” prefix and “ms” prefix was added before sub field number. To convert BD which has got one or more full texts into SOAP we will use combination of SOAP and MIME. For example the DB which we see below has got one PDF formatted full text which is called spring\_in\_action.pdf. In this case at first we convert only MARC fields and theirs data into SOAP, the end of SOAP message we include full text in MIME format.



ISSN: 2350-0328

## International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 6 , June 2020

```
<?xml version="1.0"?>
<soap:Envelope xmlns:soap="http://www.w3.org/2001/12/soap-envelope"
soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding">

<soap:Body xmlns:m="http://www.natlib.uz/carmat">
.....
</soap:Body>
</soap:Envelope>
-MIME_boundary
Content-Type: application/pdf
Content-Transfer-Encoding: 8 bit
Content-ID: <spring_in_action.pdf@carmat.uz>

...binary PDF ...
--MIME_boundary--
```

The BD have got different types full text, and it is difficult to find suitable format when we convert full text into MIME format. Because if we use wrong file format, sometimes the structure of SOAP will be failed. For example full text was in the HTML format and we include into SOAP it as TXT format. In this case user of this SOAP can not open it. To solve this problem we should use BASE64 encoding.

Base64 is a generic term for any number of similar encoding schemes that encode binary data by treating it numerically and translating it into a base 64 representation. The Base64 term originates from a specific MIME content transfer encoding. Base64 encoding schemes are commonly used when there is a need to encode binary data that needs be stored and transferred over media that are designed to deal with textual data. This is to ensure that the data remains intact without modification during transport. Base64 is used commonly in a number of applications including email via MIME, and storing complex data in XML[3].

The particular choice of characters to make up the 64 characters required for base varies between implementations. The general rule is to choose a set of 64 characters that is both part of a subset common to most encodings, and also printable. This combination leaves the data unlikely to be modified in transit through systems, such as email, which were traditionally not 8-bit clean[1]. For example, MIME's Base64 implementation uses A–Z, a–z, and 0–9 for the first 62 values. Other variations, usually derived from Base64, that share this property but differ in the symbols chosen for the last two values; an example is UTF-7.

After converting BD's full text into BASE64 text will use base64 word as Content-Transfer-Encoding of MIME format value.

```
<?xml version="1.0"?>
<soap:Envelope xmlns:soap="http://www.w3.org/2001/12/soap-envelope"
soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding">
<soap:Body xmlns:m="http://www.natlib.uz/carmat">
.....
</soap:Body>
</soap:Envelope>
-MIME_boundary
Content-Type: application/pdf
```



ISSN: 2350-0328

## International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 6 , June 2020

Content-Transfer-Encoding: base64

Content-ID: <spring\_in\_action.pdf@carmat.uz>

...base64 word of PDF ...

--MIME\_boundary--

### IV. ACKNOWLEDGEMENT

This research project is used in System ARMAT.

### REFERENCES

- [1] SOAP // <http://en.wikipedia.org/wiki/SOAP> (дата обращения 15.11.2019).
- [2] MARC 21 formats // [http://en.wikipedia.org/wiki/MARC\\_standards#MARBI](http://en.wikipedia.org/wiki/MARC_standards#MARBI) (дата обращения 15.11.2019).
- [3] *The Base16, Base32, and Base64 Data Encodings*. IETF. October 2006. RFC 4648. <http://tools.ietf.org/html/rfc4648>. Retrieved March 18, 2010.

### AUTHOR'S BIOGRAPHY



**Jasurbek Atadjanov** is head of billing system development department at Uztelecom. He has more than 17 years of experience in software development. In 2012 Jasur successfully did his Ph.D. work about developing an Information system for corporate library networks. In Uztelecom he developed a billing system for the telecommunication area. Nowadays it is used for all regional branches of the company for more than 1.5 mln clients.