Foreword (March 18th, 2003)

By its very success, the document "OpenURL Syntax Description, version OpenURL/1.0f - 2000-05-16" had become a de-facto standard by the time NISO Committee AX was formed in February 2001.

Early in the standardization process, NISO Committee AX decided to redevelop all concepts from the ground up and to put in place a solid foundation for further generalization and future growth. Simultaneously, the Committee decided that any new standard had to provide a gradual upgrade path from the existing OpenURL. In fact, the Standard that eventually emerged contains several provisions to facilitate the upgrade.

The Standard formalizes the OpenURL as it existed since May 16th, 2000, into an actual standard and assigns it the name "NISO OpenURL Version 0.1" (long version) or "OpenURL 0.1" (short version). When the Standard refers to OpenURL 0.1, it refers to the document that follows.

OpenURL Generator | OpenURL syntax description | CookiePusher document | Feedback

OPENURL SYNTAX DESCRIPTION

[Draft version, open for public comment. Please mail feedback]

Authors: Herbert Van de Sompel - Cornell University ; Patrick Hochstenbach - Ghent University ; Oren Beit-Arie - Ex Libris (USA), Inc.

This version: OpenURL/1.0f - 2000-05-16

INTRODUCTION

In order to allow for the delivery of context-sensitive services via an SFX-inspired framework, information resources must achieve the following:

1. *Implementation of a technique to make the resource understand the difference between a user that has access to a service component that can deliver context-sensitive services; and a user that does not. A pragmatic approach to this problem is described in the CookiePusher document.*

2. *For users with access to a service component, provide an OpenURL for each metadata-object. This document describes the OpenURL.*

In order to enable the delivery of context-sensitive services for -- initially bibliographic -- metadata, information providers are invited to add an OpenURL to the metadata, when it is being displayed as a result of a search/browse in their information systems. The OpenURL is designed to enable the transfer of the metadata from the information service to a service component that can provide context-sensitive services for the transferred metadata.

In order to avoid the display of the OpenURL for users working from an environment that does not have such a service component, information providers can use several techniques. The use of the CookiePusher -- that is available as a freeware tool (see CookiePusher document) -- will most probably be the easiest way for information providers to achieve this. The CookiePusher informs the information provider about the fact that a user has access to a service component. It also tells the information provider where the service component is located (see BASE-URL, below). But there are many alternative ways in which an information provider can address this problem, and the decision on how to tackle the issue will be his.

This document describes the syntax of the OpenURL for bibliographic metadata. This document is open to the public. As such, all interested parties can implement the OpenURL as part of the output of their information systems. In the same way, interested parties can create service components that can take OpenURLs as input.

0. Preliminary remarks

HTTP POST and GET

The OpenURL syntax description that is provided from item (1) onwards, uses an HTTP GET request format. However, the same syntax can also be used in an HTTP POST format. Some comments that relate to this:

- It must be understood that an OpenURL using the HTTP GET request format of a length that is higher than 255 characters may not function successfully in all circumstances. With this regard, RFC2616 mentions: "Servers ought to be cautious about depending on URI lengths above 255 bytes, because some older client or proxy implementations might not properly support these lengths." There are no such limits for a HTTP POST request format.
- While it may not be a fundamental problem for companies in the information industry to use a HTTP POST

in stead of an HTTP GET format for the OpenURL, it must be understood that the usage of a GET request format may be easier to use for an individual who wants to include an OpenURL in an HTML page he is authoring.

Character set

The OpenURL follows the URI specs (see http://www.ietf.org/rfc/rfc2396.txt). The syntax rules for URIs restrict a few characters to special roles in certain contexts and require that if these characters are used in any other way that they be Escape encoded as a percent sign followed by the character code in hexadecimal (see http://www.ietf.org/rfc/rfc2279.txt).

- The **BASE-URL** mentioned under (1) corresponds with the **<authority><path>** component of the URI specification and must comply with the rules regarding their reserved characters.
- The **QUERY** part mentioned under (1) corresponds to the **query** component of the URI specification. The declarations shown below will be used in the OpenURL syntax description, to describe the validity of characters in the different components of the **query** part of the OpenURL.

VCHAR ::= ALPHANUM | MARK | ESCAPED

ALPHANUM ::= ALPHA | DIGIT

ALPHA ::= LOWALPHA | UPALPHA

LOWALPHA ::= 'a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' | 'i' | 'j' | 'k' | 'l' | 'm' | 'n' | 'o' | 'p' | 'q' | 'r' | 's' | 't' | 'u' | 'v' | 'w' | 'x' | 'y' | 'z'

UPALPHA ::= 'A' | 'B' | 'C' | 'D' | 'E' | 'F' | 'G' | 'H' | 'I' | 'J' | 'K' | 'L' | 'M' | 'N' | 'O' | 'P' | 'Q' | 'R' | 'S' | 'T' | 'U' | 'V' | 'W' | 'X' | 'Y' | 'Z'

DIGIT ::= '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'

MARK ::= '-' | '_' | '.' | '!' | '~' | '*' | '''' | '(' | ')'

ESCAPED ::= '%' HEX HEX

HEX ::= digit | 'A' | 'B' | 'C' | 'D' | 'E' | 'F' | 'a' | 'b' | 'c' | 'd' | 'e' | 'f'

1. OpenURL

The OpenURL syntax is described here as an HTTP GET request of the form:

OpenURL ::= BASE-URL '?' QUERY

QUERY ::= DESCRIPTION ('&&' DESCRIPTION)

- *BASE-URL is the URL of a service-component that can take an OpenURL as input.*

- *DESCRIPTION describes the origin of the transported metadata-object as well as the metadata-object itself.*

- *If multiple objects are transported over the OpenURL, their DESCRIPTION must be delimited by two ampersands.*

Example:

- A **BASE-URL** could be `http://sfxserver.uni.edu/sfxmenu`
- The **BASE-URL** will depend on the user (or its institution) and can -- for instance -- become known to the information provider via the CookiePusher mechanism.

2. DESCRIPTION

DESCRIPTION ::= (ORIGIN-DESCRIPTION '&')? OBJECT-DESCRIPTION | OBJECT-DESCRIPTION ('&' ORIGIN-DESCRIPTION)?

- *OBJECT-DESCRIPTION contains information about the metadata-object transported in the OpenURL.*

- *ORIGIN-DESCRIPTION contains information about the information system where the transported metadata-object originates. It describes the system that inserts the OpenURL.*

- *The OpenURL must transport at least one object. As such the OpenURL must contain at least one OBJECT-DESCRIPTION.*

- *The order in which OBJECT-DESCRIPTION and ORIGIN-DESCRIPTION are provided is not significant.*

3. ORIGIN-DESCRIPTION

ORIGIN-DESCRIPTION ::= sid '=' VendorID ':' DatabaseID

VendorID ::= (ALPHANUM)+

DatabaseID ::= (ALPHANUM | ESCAPED)+

- *The **ORIGIN-DESCRIPTION** consists of the **sid** tag-name (**s**ervice **id**entifier) and a corresponding tag-value. This tag-value consists of two parts that are separated by a colon. The part before the colon is the identifier of the vendor of the information service where the metadata originates. The part of the tag-value following the colon is the identifier of the database within the vendor's information service where the metadata originates. The colon is provided 'as is', meaning in a non Escape encoded form.*
- *It is <u>highly</u> recommended to provide an **ORIGIN-DESCRIPTION**. If the **OBJECT-DESCRIPTION** contains a **LOCAL-IDENTIFIER-ZONE** (see 7.) then the provision of **ORIGIN-DESCRIPTION** is <u>mandatory</u>.*

Examples of **ORIGIN-DESCRIPTION** are:

- sid=Ovid:Medline
- sid=ERL:BX4
- sid=EBSCO:MFA

4. OBJECT-DESCRIPTION

OBJECT-DESCRIPTION ::= ZONE ('&' ZONE) *

ZONE ::= (GLOBAL-IDENTIFIER-ZONE | OBJECT-METADATA-ZONE | LOCAL-IDENTIFIER-ZONE)

The tag-names and corresponding tag-values that can be provided in OBJECT-DESCRIPTION resort under one of three **ZONE**(s):

- ○ The **GLOBAL-IDENTIFIER-ZONE**;
- ○ The **OBJECT-METADATA-ZONE;**
- ○ The **LOCAL-IDENTIFIER-ZONE**.

- ○ *All **ZONE**(s)are optional, but at least one of the three must be provided.*

- ○ *Each zone can only occur once in an **OBJECT-DESCRIPTION** for a transported metadata-object.*

- ○ *The choice regarding which **ZONE**(s) to provide will depend on the information system for which the OpenURL is implemented.*

- ○ *The order in which the **ZONE**(s) occur is not significant.*

5. GLOBAL-IDENTIFIER-ZONE

GLOBAL-IDENTIFIER-ZONE ::= 'id' '=' GLOBAL-NAMESPACE ':' GLOBAL-IDENTIFIER ('&' 'id' '=' GLOBAL-NAMESPACE ':' GLOBAL-IDENTIFIER)*

GLOBAL-NAMESPACE ::= ('doi' | 'pmid' | 'bibcode' | 'oai')

GLOBAL-IDENTIFIER ::= VCHAR+

The **GLOBAL-IDENTIFIER-ZONE** contains identifiers of global namespaces and the corresponding identifiers of the transported object within these global namespaces. Identifiers that only have significance in local namespaces -- such as the identifier of a record in an institutional implementation of an A&I database -- do not fit into this zone. They belong in the **LOCAL-IDENTIFIER-ZONE.**

- ○ *The **GLOBAL-IDENTIFIER-ZONE** consists of the **id** tag-name (**id**entifier) and a corresponding tag-value. This tag-value consists of two parts that are separated by a colon. The part before the colon is the identifier of the global namespace. The part of the tag-value following the colon is the identifier of the object within the global namespace.*
- ○ *The colon is provided 'as is', meaning in a non Escape encoded form.*

- ○ *More than one global identifier can be provided in the OpenURL.*
- ○ *Currently defined global namespace-identifiers are:*

  - ■ **doi** *: digital object identifier*
  - ■ **pmid** *: PubMed identifier*
  - ■ **bibcode** *: identifier used in Astrophysics Data System*
  - ■ **oai** *: identifier used in the Open Archives initiative*

Example:

- A **GLOBAL-IDENTIFIER-ZONE** can be: `id=doi:123/345678&id=pmid:202123`
- A valid OpenURL -- before the mandatory Escape encoding -- is:
  `http://sfxserver.uni.edu/sfxmenu?id=doi:123/345678&id=pmid:202123`
  This OpenURL transports two global identifiers that uniquely define the same metadata-object.
- The corresponding Escape encoded OpenURL is: `http://sfxserver.uni.edu/sfxmenu?id=doi:123%2F345678&id=pmid:202123`
- A valid OpenURL -- before the mandatory Escape encoding -- for a preprint that resides in an archive that complies with the Santa Fe Convention of the Open Archives initiative is:
  `http://sfxserver.uni.edu/sfxmenu?id=oai:arXiv:physics/0003005`

- The corresponding Escape encoded OpenURL is
  `http://sfxserver.uni.edu/sfxmenu?id=oai%3AarXiv%3Aphysics%2F0003005`

6. OBJECT-METADATA-ZONE

OBJECT-METADATA-ZONE ::= META-TAG '=' META-VALUE (& META-TAG '=' META-VALUE) *

META-TAG ::= ('genre' | 'aulast' | 'aufirst' | 'auinit' | 'auinit1' | 'auinitm' | 'coden' | 'issn' | 'eissn' | 'isbn' | 'title' | 'stitle' | 'atitle' | 'volume' | 'part' | 'issue' | 'spage' | 'epage' | 'pages' | 'artnum' | 'sici' | 'bici' | 'ssn' | 'quarter' | 'date')

META-VALUE ::= VCHAR+

The **OBJECT-METADATA-ZONE** is used for the provision of metadata elements of the transported metadata-object in a format that is shared by all OpenURLs. If for some reason metadata elements can not be described in this common format, they can still be included in the **PRIVATE-IDENTIFIER-ZONE.**

- **Table 1** shows a list of currently supported **META-TAG**s and a description of their meaning.
- **Table 2** shows the usage of **META-TAG**s in relation to the **genre** of the transported object.

Example:

- An **OBJECT-METADATA-ZONE** can be :
  `issn=1234-5678&date=1998&volume=12&issue=2&spage=134`
- A valid OpenURL can be : `http://sfxserver.uni.edu/sfxmenu?issn=1234-5678&date=1998&volume=12&issue=2&spage=134` . Note that the "-" in the **issn** tag-value is part of the **VCHAR** set and as such does not need to be Escape encoded.

7. LOCAL-IDENTIFIER-ZONE

LOCAL-IDENTIFIER-ZONE ::= 'pid' '=' VCHAR+

The **LOCAL-IDENTIFIER-ZONE** is introduced in order to allow for the transportation of metadata in formats that are specific to the originating information system, and that can not be expressed in the standardized syntax proposed for the **OBJECT-METADATA-ZONE.**

- The **LOCAL-IDENTIFIER-ZONE** consits of a **pid** (**p**rivate **id**entifier) tag-name and a corresponding tag-value. The syntax of the tag-value is completely defined by the information provider.

- If a **LOCAL-IDENTIFIER-ZONE** is used, then the provision of **ORIGIN-DESCRIPTION** (see 3.) is mandatory.

- The **LOCAL-IDENTIFIER-ZONE** must be Escape encoded as a whole, meaning that -- for instance -- also parameter-names defined by the information providers must be Escape encoded.

Example:

- A **LOCAL-IDENTIFIER-ZONE** can be: `pid=<author>Smith, Paul ; Klein, Calvin</author>&<yr>98</yr>`
- An OpenURL containing the above **LOCAL-IDENTIFIER-ZONE** -- before the mandatory Escape encoding -- would be :
  `http://sfxserver.uni.edu/sfxmenu?`

```
sid=EBSCO:MFA&id=pmid:203456&pid<author>Smith, Paul ; Klein,
Calvin</author>&<yr>98</yr>
```

- The corresponding encoded OpenURL is:
  ```
  http://sfxserver.uni.edu/sfxmenu?sid=EBSCO:MFA&
  id=pmid:203456&pid=%3Cauthor%3ESmith%2C%20Paul%20%3B%20Klein%2C%20Calvin%3C%
  2Fauthor%3E&%3Cyr%3E98%2F1%3C%2Fyr%3E.
  ```
  As can be seen, the **pid** value is encoded as a whole.

- Because the following OpenURL -- shown before the mandatory Escape encoding -- contains a **pid** without a **sid**, it is invalid:
  ```
  http://sfxserver.uni.edu/sfxmenu?id=pmid:203456&pid<author>Smith, Paul ;
  Klein, Calvin</author>&<yr>98</yr> .
  ```

META-TAG	value	description
genre	bundles:	
	journal	a journal, volume of a journal, issue of a journal
	book	a book
	conference	a publication bundling proceedings of a conference
	individual items:	
	article	a journal article
	preprint	a preprint
	proceeding	a conference proceeding
	bookitem	an item that is part of a book
aulast		A string with the <u>first</u> author's last name
aufirst		A string with the <u>first</u> author's first name
auinit		A string with the <u>first</u> author's first and middle initials
auinit1		A string with the <u>first</u> author's first initial
auinitm		A string with the <u>first</u> author's middle initials
issn		An ISSN number
eissn		An electronic ISSN number
coden		A CODEN
isbn		An ISBN number
sici		A SICI of a journal article, volume or issue. Compliant with ANSI/NISO Z39.56-1996 Version 2 (see http://sunsite.berkeley.edu/SICI/)

bici		A BICI for a section of a book, to which an ISBN has been assigned. Compliant with http://www.niso.org/bici.html
title		The title of a bundle (journal, book, conference)
stitle		The abbreviated title of a bundle
atitle		The title of an individual item (article, preprint, conference proceeding, part of a book)
volume		The volume of a bundle
part		The part of a bundle
issue		The issue of a bundle
spage		The start page of an individual item in a bundle
epage		The end page of an individual item in a bundle
pages		Pages covered by an individual item in a bundle. The format of this field is `spage-epage`
artnum		The number of an individual item, in cases where there are no pages available.
date	YYYY-MM-DD YYYY-MM YYYY	The publication date of the item or bundle encoded in the "Complete date" variant of ISO8601 (see http://www.w3.org/TR/NOTE-datetime). This format is YYYY-MM-DD where YYYY is the four-digit year, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 28 or 29 or 30 or 31, depending on length of the month and whether it is a leap year.
ssn	winter | spring | summer | fall	The season of publication
quarter	1 | 2 | 3 | 4	The quarter of publication

Table 1 : META-TAGs and description of their meaning

	genre						
	individual items				bundles		
	article	preprint	proceeding	bookitem	book	journal	conference
aulast	X	X	X	X	X	-	X
aufirst	X	X	X	X	X	-	X
auinit	X	X	X	X	X	-	X
auinit1	X	X	X	X	X	-	X
auinitm	X	X	X	X	X	-	X

`issn`	X	-	X	-	-	X	X
`eissn`	X	-	X	-	-	X	X
`coden`	X	-	X	-	-	X	X
`isbn`	-	-	X	X	X	-	X
`sici`	X	-	X	-	-	X	X
`bici`	-	-	X	X	-	-	-
`title`	X	-	X	X	X	X	X
`stitle`	X	-	X	X	X	X	X
`atitle`	X	X	X	X	-	-	-
`volume`	X	-	X	X	X	X	X
`part`	X	-	X	X	X	X	X
`issue`	X	-	X	-	-	X	X
`spage`	X	X	X	X	-	-	-
`epage`	X	X	X	X	-	-	-
`pages`	X	X	X	X	-	-	-
`artnum`	X	X	X	X	-	-	-
`date`	X	X	X	X	X	X	X
`ssn`	X	X	X	X	X	X	X
`quarter`	X	X	X	X	X	X	X

Table 2 : META-TAGs and how they relate to genres

History

2000-05-16 : Made changes to address ambiguity regarding Escape encoding of the different components of the OpenURL.

2000-05-12 : Added a section with regard to HTTP GET and POST. Added the names of the authors of the OpenURL document.

2000-05-02 : Selective release of OpenURL specs to a community of experts in reference linking

Currently under discussion

Addition of a date-type tag to accomodate for difference in publication date of print and electronic versions. Such a tag is used in the CrossRef DTD.

Addition of OpenURL version number in the syntax.

Meaning of the genre tag. In OpenURL, the tag-value of the genre tag corresponds with the type of the object that is described in the OpenURL. In CrossRef, the genre tag refers to the type of the object itself.

Acknowledgements

Many thanks to Tony Hammond at Academic Press for valuable input.

This page last updated on: 2 November, 2001