

Grant Agreement ECP-2007-DILI-527003

ARROW

**Specification
for metadata messaging formats**

Deliverable number	<i>D4.3.2</i>
Dissemination level	<i>Public</i>
Delivery date	<i>May 31st, 2010</i>
Status	<i>Final</i>
Author(s)	<i>EDItEUR, BnF</i>



eContentplus

This project is funded under the *eContentplus* programme¹,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

Content of deliverable D4.3.2

This deliverable contains all the documentation for version 0.2 of the ARROW Message Descriptions developed up to April 30th, in order to support the implementation of the Beta version of the ARROW prototype.

The materials delivered are as follows:

- This document giving an overview of the ARROW message formats.
- A PDF document containing detailed descriptions of each message (*D4.3.2_ANNEX_I_20100701_ARROW Message Definitions version 0_25.pdf*)
- A glossary of terms used (*D4.3.2_ANNEX_II_20100531_ARROW Glossary of terms*)
- A zipped file containing XML schemas, code lists and HTML documentation for each message (*ARROW message schemas v0.2 20100430.zip*), see below.

The schemas in the zipped file have been made available (a) for each query-and-response message pair, (b) for a number of common components used across the message set and (c) for an internal reference model known as the reference schema. Similarly, the HTML documentation can be accessed either for individual messages, for common components or for the reference schema, in each case delivering structured and pictorial representations of each message element.

TABLE OF CONTENTS

1. Introduction.....	4
2. ARROW: objectives & participants.....	4
3. Message types & choreography.....	5
4. Message structure.....	7
5. Related resources.....	8
6. Detailed message descriptions.....	8

ARROW: message format overview

Version 0.21, 30 May 2010

1. Introduction

This document describes the messages in ONIX format that EDItEUR has developed to support the ARROW project based on the business requirements and feedback of the ARROW technical working group.¹

The messages are designed to be used by various players in the ARROW workflow and to cover the basic stages in the process, from an initial library request through to the grant or denial of a license. ARROW itself is described extensively elsewhere: see <http://www.arrow-net.eu/> for more information.

EDItEUR's role is to support ARROW by creating and maintaining a suite of robust and extensible messages to allow for structured communication between ARROW participants. The messages have been developed in ONIX format, building upon extensive work by EDItEUR in a number of related areas, including ONIX for Books (book product metadata), ONIX-PL (publications licenses) and ONIX for ISTC Registration (registration of works and their identifiers).

The messages described here are currently in use for the Beta pilot phase of the ARROW project. Going forward, the messages will continue to be supported by EDItEUR and will evolve as ARROW business requirements develop. Since the ARROW messages are closely interrelated with, and in some cases derived from, other ONIX formats they will continue to utilize a number of shared ONIX code lists and any updates or enhancements therein.

2. ARROW: objectives and participants

To provide context for the messaging requirements, ARROW's main aims may be summarized as follows:

- To help identify the works represented by printed books that libraries wish to digitize and make available online to their clients; and to carry out similar steps for books that have already been digitized.
- To help to clarify the copyright status of a textual work in order to determine whether it is in public domain or still in copyright, whether in print or out of print.
- To discover the rights holders for these works and facilitate library and other user requests to obtain licenses, where necessary, for digitization and/or usage of digitized copies of the books.
- To support the establishment of a Registry of Orphan Works, covering those works for which no rights holders can be discovered after diligent search has been undertaken.

A number of organizations across Europe are collaborating to build the ARROW service. In terms of messaging partners, these participants and their roles may be briefly characterized as follows:

- For the time being, libraries are ARROW's primary "customers", as owners of the books and requesters of permissions and licenses.
- A central ARROW infrastructure has been created to manage access to the ARROW workflow, its shared services and its information base.
- The European Library (TEL) coordinates access to the resources of European national libraries and the Virtual International Authority File (VIAF) and provides resource identification and clustering services to ARROW.
- Books in Print agencies match incoming requests against their own records and advise ARROW of any matching ISBNs as well as returning ONIX product records containing status, availability and other metadata.

¹ The ARROW technical working group was set up in October 2009 and consist of: AIE as project coordinator, BnF as the leader of WP4 (Interoperability), MVB as the leader of WP5 (Design of system architecture) and technical expert for the BiP domain, Cineca as the leader of WP6 (Set up of rights information infrastructure), The European Library (TEL) as technical expert for the Library domain, ALCS as technical expert for the RRO domain. Whenever necessary collaboration with standard governing organisations has been established (ISBN International Agency, ISTC international Agency for example).

- Reproduction Rights Organizations (RROs) review the consolidated requests from the libraries and the information gathered throughout the ARROW process. On this basis the RROs then proceed to grant or deny licenses, as appropriate, as well as offering further advice to ARROW and the libraries on how best to bring requests to complete resolution.

3. Message types & choreography

The ARROW messages have been designed as a series of “query and response” pairs. That is to say, the requesting party sends a precisely formulated request and then expects a similarly structured response from the party that has been queried. This methodology can be used for real-time or synchronous interactions between systems linked by web services as well as for asynchronous interactions where it is expected that there may be some delay or intervening processes before the response is sent.

In the current pilot release, Version 0.2, fourteen messages (seven request and response pairs) have so far been defined and deployed. It is anticipated that further messages may be required later, in particular when the proposed Registry of Orphan Works has been established.

The sender and addressee (receiver) of each message are defined by an ARROW workflow that can be, for the time being, visualized as essentially sequential in nature.

Starting with the initial request from a library, the resulting “transaction” passes through stages of resource identification, work and manifestation clustering and the identification of related resources, before submission to an RRO for a licensing decision. For the time being, the ARROW workflow runs on a national basis based on the country of publication of the target resource. A number of facilities are built in to allow the library or other interested party to review the matching or clustering undertaken, and to return details of any decisions. The pilot process workflow concludes with the RRO sending its considered response to ARROW: this may involve the grant or denial of a license and/or the provision of additional information to help the library bring the request to complete resolution.

The table on the next page summarizes the messages in Version 0.2, the name and business purpose of each, and the sender and addressee for each exchange. Note that the terminology used there is further explained in a companion document, the *ARROW Glossary (D4.3.2_ANNEX_II_20100531_ARROWGlossary of terms)*.

No.	Sender	Addressee	Message name* and purpose
M1Q	Library	ARROW	InitialResourceAndUsageRequest Asks ARROW to identify a "target resource" held by the library, assist in locating rightsholder(s) for the corresponding work, and convey details of the usages for which license or other permissions are sought
M1R	ARROW	Library	InitialResourceAndUsageResponse Acknowledges the library's request and assigns a persistent transaction identifier to the request to support further operations or inquiries
M2Q	ARROW	TEL	ResourceIdentificationRequest Asks a "reference source" (initially TEL) to identify or confirm the identity of a published "target resource" held by a library
M2R	TEL	ARROW	ResourceIdentificationResponse Communicates the results of attempts to match details of a library's "target resource" with the reference source's own records
M3Q	ARROW	Library	ManifestationMatchingReviewRequest Asks the library to review the results of attempts by a reference source (initially TEL) to match details of the library's "target resource" with the reference source's own records
M3R	Library	ARROW	ManifestationMatchingReviewResponse Communicates the results of the library's review of matches submitted by a reference source (initially TEL) between the library's "target resource" and the reference source's own records
M4Q	ARROW	TEL	ClusterCreationRequest Conveys details of a particular manifestation and asks a reference source (initially TEL) to return details of one or more clusters of other manifestations/works from its own records that appear to be related to that manifestation
M4R	TEL	ARROW	ClusterCreationResponse Communicates the results of attempts to identify clusters of manifestations/works from the reference source's own records, based upon the original manifestation originally submitted
M5Q	ARROW	Library	ClusterReviewRequest Asks the library to review the results of attempts by a reference source (initially TEL) to identify clusters of manifestations/works from the reference source's own records, based upon the manifestation originally submitted
M5R	Library	ARROW	ClusterReviewResponse Communicates the results of the library's review of clusters submitted by a reference source (initially TEL) based upon manifestations related to the library's "target resource" and the reference source's own records
M6Q	ARROW	BIP	RelatedBooksInPrintRequest Conveys details of a particular manifestation and asks the BIP to return details of any manifestations from its own records that appear to be related to that manifestation, together with information on the publishing status and availability of each one
M6R	BIP	ARROW	RelatedBooksInPrintResponse Communicates details of any manifestations (or "related ISBNs") identified by the BIP's matching between a particular manifestation and its own records, together with information on the publishing status and availability of each one
M7Q	ARROW	RRO	FormalLicenseRequest Identifies a "target resource" for which a library requests a formal license or other permissions. It specifies the usage permissions sought and presents supporting information gathered through the ARROW process, including details of any identified related works or manifestations and the apparent publishing status or availability of each one
M7R	RRO	ARROW	LicenseProposalOrRefusal Conveys the RRO's decision or other responses concerning the original request from a library. Responses may include the proposal or refusal of a license, advice that a license is unnecessary, and a range of other advice

*Each name is preceded by the word 'ARROW' in the message itself.

4. Message structure

The suite of messages developed for ARROW draws extensively upon existing formats from the ONIX family of standards, created and maintained by EDItEUR. This has the advantage that a great many concepts have been straightforwardly derived from published ONIX work in several areas closely related to ARROW's needs. These include:

- ONIX for Books, for bibliographic descriptions and metadata, as well as supply chain status and availability.
- ONIX-PL (Publications License) and ONIX for License Terms, for terminology and a structured vocabulary to describe licenses, rights, usage, etc.
- ONIX for ISTC Registration, for some concepts related to works, their identifiers and their relation to other resources. (Although note that the ONIX for ISTC Registration messages themselves will be used unchanged by ARROW participants for formal ISTC registration.)

At the same time, the ONIX methodology is extensible and a number of concepts and controlled values have been introduced and will be maintained in a separate ARROW namespace to cater for requirements that are specific to ARROW.

All ONIX messages, including those for ARROW, are expressed in XML. This provides platform- and software-independence for those implementing the messages, together with intuitive message structuring and ease of integration with web-enabled and browser-based applications.

ARROW necessarily operates at the interface between the library domain (where variants of MARC are used extensively in cataloguing and bibliographic records) and the book-publishing domain (where usage of ONIX for Books records is widespread, particularly in Europe and North America). For this reason, the ARROW messages can carry descriptive "payloads" expressed in a variety of formats, including MARC21, UNIMARC, ONIX for Books 2.1 & 3.0, and an amended subset of ONIX for Books known as the ARROW ShortDescription.

Whatever its specific purpose, each ARROW message has the same high-level structure, as illustrated in this example:

	Structure	Explanation	Cardinality
1	<ARROWInitialResourceAndUsageRequest version="0.2">	An opening tag, specifying the type and purpose of the message, with an attribute showing the version number	
2	<Header>	A block of administrative information, identifying sender and receiver, as well as the date, time and number of the message	1
3	<ARROWTransaction>	One or more "transactions", which carry the "payload" of the message	1-n
4	<Summary>	An optional block of control totals (where relevant) to check the integrity of message creation	0-1
5	</ARROWInitialResourceAndUsageRequest>	A closing tag, signifying the end of the message	

Messages themselves are composed of a sequence of defined data elements, each of which is fully defined in the relevant detailed message description. Some elements are "simple" and just consist of opening and closing tags surrounding the element value, like this:

```
<ARROWTransactionID>1234567</ARROWTransactionID>
```

Others are "composites" that themselves contain several elements, such as the <SenderIdIdentifier> composite here:

```
<SenderIdIdentifier>
  <SenderIdType>06</SenderIdType>
  <IDValue>5055123401239</IDValue>
</SenderIdIdentifier>
```

Note that message structures are often displayed in this way, with indentation and nesting used more clearly to illustrate the underlying logic; but in real XML message files, there should be no spaces, indentations or carriage returns between any of the elements!

Finally, the content of an element may in some cases be free text. But in most cases, the element is populated with one of a series of controlled values. These are held in defined code lists maintained by EDItEUR, supplemented where necessary by values requested by the ARROW team to represent ARROW-specific concepts.

The XML schemas supplied along with this documentation control and enforce the logic of the message structure, validate the use of the correct controlled values and check that the message parses correctly, either before sending or upon receipt.

5. Related resources

Further information about a number of the resources or related standards mentioned here may be found on the EDItEUR website at <http://www.editeur.org/>

Specifically, the following relevant ONIX standards are fully documented:

- ONIX for Books, version 3.0: <http://www.editeur.org/12/About-Release-3.0/>
- ONIX for Books code list 11: <http://www.editeur.org/93/Release-3.0-Downloads/#Codes%2010>
- ONIX for Publications Licenses, version 1.0: <http://www.editeur.org/21/ONIX-PL/>
- ONIX for Licensing Terms: <http://www.editeur.org/85/Overview/>
- ONIX for ISTC Registration, version 1.0: <http://www.editeur.org/106/ONIX-ISTC-Registration-Format/>

6. Detailed message descriptions

During the pilot phases of the ARROW project, requirements have been progressively modified and fine tuned so that they more accurately reflect the evolving ARROW workflow. Detailed descriptions of the messages created to date can be found in the companion PDF document (*D4.3.2_ANNEX_I_20100701_ARROW Message Definitions version 0_25.pdf*).

For technical implementers, comprehensive documentation is also available in the set of HTML and image files delivered along with the XSD schemas. These provide click-through descriptions of each message pair and also of an overall structure known as a reference schema.