# ONIX for Preservation Holdings
## DRAFT Format Overview
## Version 0.21

March 2012

---

EDItEUR invites comments on this draft specification and the associated XML schema.
Please send comments and suggestions for improvement to info@editeur.org.

---

**ONIX-PH Version 0.21: Summary & Change Control**
This draft document has been prepared together with the EDItEUR Working Group on preservation messaging.

The document has two main parts.  The first part, covering sections 1–7, is an introduction outlining the purpose of ONIX for Preservation Holdings, describing how it has been developed and documented, and explaining some of the message features and terminology used.  The second part of the document presents an extended tabular description of the structure of the message.

Significant changes since the August 2011 (version 0.2) draft are as follows:

- The version number and issue date have been changed.
- Detailed change control information, summarizing recent decisions within the working group, has been removed.
- Information related to the associated ONIX for Serials controlled codelists has been integrated into the tabular description.

**EDItEUR**
39-41 North Road
LONDON N7 9DP
UK

Telephone +44 (0)20 7503 6418
Fax +44 (0)20 7503 6418
http://www.editeur.org

EDItEUR is the international group that coordinates the development and promotion of
standards for electronic commerce in the books and serials sectors.

## 1. Introduction

This document has two main parts.  The first part, covering sections 1–7, is a descriptive overview, outlining the purpose of ONIX for Preservation Holdings, describing how it has been developed and documented, and explaining some of the message features and terminology used.  The second part of the document presents an extended tabular description of the structure of the message.

## 2. ONIX for Preservation Holdings

The ONIX for Preservation Holdings format (ONIX-PH) is designed for structured, machine-to-machine communications about online resources that have been preserved or committed for preservation. Although initially formulated with online journals in mind, the format can also accommodate information about other kinds of resource, such as e-books, e-book collections or online databases, as these gain importance in the library supply chain.

ONIX-PH has been developed in close cooperation between EDItEUR and a specialized preservation messaging Working Group with representatives from preservation agencies, publishers, the PEPRS project group, and other intermediaries.  PEPRS has established a prototype e-journal preservation registry service known as The Keepers Service. Communications between preservation agencies and the prototype registry define the initial business use cases for ONIX-PH.

Since the major part of the message "payload" deals with statements of online holdings, ONIX-PH has drawn heavily on earlier work that defined the ONIX for Serials SOH (serials online holdings) format.  The current, version 0.3 draft ONIX-PH message is intended to support pilot exchanges between Working Group partners.  On successful conclusion of the pilots, necessary modifications will be incorporated into the first formal version 1.0 release. Thereafter every effort will be made to maintain backwards compatibility between successive releases.

## 3. Message structure

In common with all ONIX formats, the data elements that comprise ONIX-PH are expressed in XML and controlled by an XML schema. The tabular description later in this document describes each element of the message, as well as the controlled-value codelists utilized in many cases.  The tabular description also specifies the "cardinality" expected by the schema for each element – in simple terms, whether a particular element is optional or mandatory and whether it may appear only once or many times.

## 4. Relation to ONIX for Serials SOH

ONIX-PH may be viewed as a "sibling" of the ONIX for Serials SOH format, specialized for preservation applications.  First developed as a collaborative effort by EDItEUR and a NISO Joint Working Party, the SOH format supports a number of business cases that require communication of online holdings information. Readers who require a more general treatment of online holdings may find it useful to review SOH. The latest release of SOH, version 1.2, was published in June 2011 and comprehensive details are available from the EDItEUR website here.

## 5.   About holdings descriptions and the Coverage Statement

There are many variations in the ways in which publishers describe and enumerate the resources that they publish. To cater for this complexity, the ONIX for Serials Coverage Statement was developed.  The Coverage Statement is not in itself a complete messaging format; rather, it provides a structured and multilevel architecture within which complex enumeration schemes can be communicated, whether for single releases or for one or several ranges of releases or issues.

The Coverage Statement is incorporated within the <PackageDetail> composite of ONIX-PH. Coverage itself is comprehensively documented here.

For e-books and continually updated resources (like reference works), there is no Coverage as such; the holdings are simply the resource itself.

## 6.   Possible future use cases and extensions

Pilot implementations of ONIX-PH will first focus upon communications of preserved holdings from holdings agencies to the prototype registry.  However, it is intended that the message architecture be able to accommodate other use cases as these are identified.  Two likely additional usage areas are under consideration.

*1. Other use cases in online resource preservation*

The format may also be used for communications from publishers originating the content to preservation agencies. Similarly, it is capable of allowing preservation agencies to communicate with individual business partners or even to broadcast statements of the resource holdings preserved.  Care has been taken to generalize the ways in which data structures are represented, so that these and other cases can be accommodated as the needs are identified.

*2.  Print resources*

The principles of ONIX-PH can in principle be extended to communicate details of preserved *print* as well as online holdings.  It is expected that a future version, with appropriate modifications, will support messages carrying print and/or electronic holdings.

## 7.   Related documentation and resources

Currently, the documentation for ONIX-PH comprises four items.  These draft materials are available from the EDItEUR website here.

The first item is this descriptive Overview, available in PDF format.  The second is an XML schema that specifies and controls the message structure and that can be used to check and parse ONIX-PH files created by participants in the pilot.  The third item is a set of the controlled code lists utilized by ONIX-PH.  The XML schema and the code lists are both provided as .xsd files.  Finally, the message is also described in HTML, generated directly from the XML schema using the proprietary <oXygen/> software.

## 8.   Acknowledgments

EDItEUR acknowledges with thanks the inputs and expertise provided by members of the Preservation Messaging Working Group.  The Group has included representatives of the following organizations:

- British Library
- E-Depot of the Koninklijke Bibliotheek
- EDINA
- EDItEUR
- Harrassowitz
- International ISSN Centre
- Library of Congress
- LOCKSS Program
- Portico/Ithaka
- Serials Solutions
- Stanford University Libraries
- Taylor & Francis
- UK LOCKSS Alliance.

**Overview of the ONIX for Preservation Holdings format**

The tables on this and subsequent pages give an overview of the ONIX for Preservation Holdings format, and show how elements are nested.  The first section, colour-coded yellow, shows the message header.  The second section, colour-coded light brown, shows the broad content of the detail section of the message.  The third section, colour-coded green, shows an expansion of the <OnlinePackage> composite from the previous page. The fourth section, colour-coded pale blue, shows an expansion of the <RelatedResourceVersion> composite.

Cardinality is indicated in the right-hand column for each element: 1 = Mandatory, not repeatable within its parent. 0-1 = Optional, not repeatable within its parent. 1-n = Mandatory, repeatable within its parent. 0-n = Optional, repeatable within its parent.

| 1 | <ONIXPreservationHoldings version="0.2"> | | | A list of online holdings preserved or committed to preservation by one or more preservation agencies | |

| 2 | **<Header>** | | | Message header | 1 |
|---|---|---|---|---|---|
| 3 | <Sender> | | | The sender of the message (coded identifier or name or both) | 1 |
| 4 | | <SenderIdentifier> | | A coded identifier of the message sender, eg a SAN or GLN | 0-n |
| 5 | | | <SenderIDType> | A code indicating the scheme from which the identifier is taken. See code list 44A for permissible values | 1 |
| 6 | | | <IDTypeName> | The name of a proprietary scheme, if applicable | 0-1 |
| 7 | | | <IDValue> | The identifier value | 1 |
| 8 | | <SenderName> | | The name of the sender organisation | 0-1 |
| 9 | | <SenderContact> | | The name of a contact person in the sender organisation | 0-1 |
| 10 | | <SenderEmail> | | An email address for the sender | 0-1 |
| 11 | <Addressee> | | | The addressee of the message (omitted in "broadcast" notifications) | 0-n |
| 12 | | <AddresseeIdentifier> | | A coded identifier of the message addressee | 0-n |
| 13 | | | <AddresseeIDType> | A code indicating the scheme from which the identifier is taken. See code list 44A for permissible values | 1 |
| 14 | | | <IDTypeName> | The name of a proprietary scheme, if applicable | 0-1 |
| 15 | | | <IDValue> | The identifier value | 1 |
| 16 | | <AddresseeName> | | The name of the addressee organisation | 0-1 |
| 17 | | <AddresseeContact> | | The name of a contact person in the addressee organisation | 0-1 |

| 18 | | <AddresseeEmail> | | An email address for the addressee | 0-1 |
|---|---|---|---|---|---|
| 19 | <MessageNumber> | | | Message sequence number | 0-1 |
| 20 | <MessageRepeat> | | | A number which distinguishes any repeat transmissions of a message | 0-1 |
| 21 | <SentDateTime> | | | The date, and optionally the time, when a message was sent | 1 |
| 22 | <MessageNote> | | | A free-text note about the contents of the message | 0-1 |
| 23 | <CompleteFile/> | | | One or other of these empty elements is mandatory, to indicate whether the message is a complete replacement file or an incremental update | 0-1 |
| 24 | <DeltaFile/> | | | | 0-1 |

| 25 | **<HoldingsList>** | | | A list of e-content preserved by a preservation agency. Repeatable, with one repeat for each preservation agency whose holdings are being transmitted in this message | 1-n |
|---|---|---|---|---|---|
| 26 | <PreservationAgency> | | | The party or agency whose preservation holdings are listed in this holdings list. | 1 |
| 27 | | <PreservationAgencyIdentifier> | | A coded identifier of a preservation agency: either <PreservationAgencyIdentifier> or <PreservationAgencyName> or both must be present. Repeatable if multiple identifiers are sent. | 0-n |
| 28 | | | <PreservationAgencyIDType> | A code indicating the scheme from which the identifier is taken. See code list 44A for permissible values | 1 |
| 29 | | | <IDTypeName> | The name of a proprietary scheme, if applicable | 0-1 |
| 30 | | | <IDValue> | The identifier value | 1 |
| 31 | | <PreservationAgencyName> | | The name of a preservation agency | 0-1 |
| 32 | | <Website> | | Details of the website(s) on which the preserved resources may be accessed | 0-n |
| 33 | | | <WebsiteRole> | A code indicating the role of a website. See code list 73. In this case the value will always be 03, "Online hosting service home page." | 1 |
| 34 | | | <WebsiteDescription> | A free-text description of the nature or function of a website | 0-1 |
| 35 | | | <WebsiteLink> | The URL for a website | 1 |
| 36 | | | <MirrorSite> | Composite: details of any mirror site(s) | 0-n |
| 37 | <HoldingsRecord> | | | Details of the preservation agency's holdings of a specified resource version. One occurrence for each distinct resource version within a <HoldingsList>. | 1-n |

| 38 | | <NotificationType> | | A code indicating whether the holdings record is new, an update or a deletion.  See code list 1A for permissible values. | 1 |
|----|---|---|---|---|---|
| 39 | | <ResourceVersion> | | Details for a specified resource version | 1 |
| 40 | | | <ResourceVersionIdentifier> | Composite: a coded identifier for the resource version, eg ISSN, ISBN. Includes <ResourceVersionIDType>, <IDTypeName>, <IDTypeValue> See code list 103S for permissible values for <ResourceVersionIDType> | 0-1 |
| 41 | | | <Title> | Composite: the title of a resource version | 0-1 |
| 42 | | | <Publisher> | Composite: the publisher of a resource version | 0-n |
| 43 | | | <MetadataProvider> | Composite: the organization that supplied the metadata for the resource included in the HoldingsList.. Identifies the organization using a structure similar to that for PreservationAgency and includes: <MetadataProviderIdentifier>     <MetadataProviderIDType>     <IDTypeName>     <IDValue> <MetadataProviderName> See code list 44A for permissible values for <MetadataProviderIDType> This composite is used only when the provider of the metadata is known.  If this composite is not supplied, nothing should be inferred; that is, nothing is known about who provided the metadata. | 0-1 |
| 44 | | | <OnlinePackage> | Composite: details of the coverage and completeness of a resource version preserved by the preservation agency. See expanded view below. | 1 |

| 1 | **\<OnlinePackage>** | | | Details of the location, coverage and completeness of a resource version preserved by the preservation agency. | 1 |
|---|---|---|---|---|---|
| 2 | \<Website> | | | If the specified resource version has its own direct URL(s), these should be included here.  If it is accessed via a generic URL for the preservation agency, which has been declared at the beginning of the holdings list, the generic URL should not be repeated at title level. | 0-n |
| 3 | | \<WebsiteRole> | | A code indicating the role of a website. See code list 73A for permissible values | 1 |
| 4 | | \<WebsiteDescription> | | A free-text description of the nature or function of a website | 0-1 |
| 5 | | \<WebsiteLink> | | The URL for a website | 1 |
| 6 | | \<MirrorSite> | | Composite: details of any mirror site(s).  Repeatable for each mirror site. | 0-n |
| 7 | | | \<WebsiteDescription> | | 0-1 |
| 8 | | | \<WebsiteLink> | | 1 |
| 9 | \<CoverageNote> | | | New in v. 1.1.  A free text note clarifying any coverage information found in \<PackageDetail> | 0-1 |
| 10 | \<PackageDetail> | | | Details of a range and completeness of content for a resource version within an online package. \<PackageDetail> is repeated if the package includes ranges with different levels of completeness and/or online formats. | 0-n |
| 11 | | \<Coverage> | | An ONIX for Serials coverage statement, showing a range of issues held by the preservation agency.  \<Coverage> is documented separately as the *ONIX for Serials: Coverage Statement.* | 1 |
| 12 | | \<MissingCoverage> | | Issues or range of issues that are known to have not been published.  This is *not* a list of issues that have not been preserved.  Identical in structure to \<Coverage>. | 0-1 |
| 13 | | \<IssueCompleteness> | | A code indicating whether all articles from the issues in the range of content in this \<PackageDetail> are available online. See code list 105S for permissible values | 0-1 |
| 14 | | \<ArticleCompleteness> | | A code indicating whether articles in the issues in the range of content in this \<PackageDetail> are held as full text, abstracts etc. See code list 106S for permissible values | 0-1 |
| 15 | | \<ContentDescription> | | A free text description of the content available online | 1 |
| 16 | | \<EpubFormat> | | A code indicating the file format(s) in which the range of online content in this \<PackageDetail> is preserved; eg HTML, PDF.  See code list 175S for permissible values. | 0-n |

| | | | | [Note: this codelist now includes values for book identifiers.] Repeatable if the holdings included in this PackageDetail are presented in multiple formats (e.g. both HTML and PDF) | |
|---|---|---|---|---|---|
| 17 | | <PreservationStatus> | | Status of preservation activities for the range of online content in this <PackageDetail> | 1 |
| 18 | | | <PreservationStatusCode> | A coded value indicating the status of preservation activities.  See code list 199 for permissible values | 1 |
| 19 | | | <DateOfStatus> | Date on which the status was recorded, in format YYYYMMDDHHMM (HHMM not required).  Note that an ONIX-PH message reflects the preservation status as of a given moment in time. | 1 |
| 20 | | <VerificationStatus> | | A code indicating whether the sending agency has verified that all issues known to have been published have been included in the <Coverage> statement.  See code list 200 for permissible values | 1 |
| 21 | <NoPackageDetail/> | | | Empty element indicating that content detail is not being provided for this online package.  This could happen, for example, when an agency has made a commitment to preserve a resource, but does not have any details about the content to be preserved. | 0-1 |
| 22 | <LicenseTerms Description> | | | A free-text statement of key license terms: effectively a placeholder for future inclusion of structured license terms information, or links to the applicable license | 0-1 |