

Protocol for Exchanging Serial Content (PESC)

Approved May 14, 2015

*A Recommended Practice of the
National Information Standards Organization*

About NISO Recommended Practices

A NISO Recommended Practice is a recommended “best practice” or “guideline” for methods, materials, or practices in order to give guidance to the user. Such documents usually represent a leading edge, exceptional model, or proven industry practice. All elements of Recommended Practices are discretionary and may be used as stated or modified by the user to meet specific needs.

This recommended practice may be revised or withdrawn at any time. For current information on the status of this publication contact the NISO office or visit the NISO website (www.niso.org).

Published by

National Information Standards Organization (NISO)
3600 Clipper Mill Road, Suite 302
Baltimore, MD 21211
www.niso.org

Copyright © 2015 by the National Information Standards Organization

All rights reserved under International and Pan-American Copyright Conventions. For noncommercial purposes only, this publication may be reproduced or transmitted in any form or by any means without prior permission in writing from the publisher, provided it is reproduced accurately, the source of the material is identified, and the NISO copyright status is acknowledged. All inquiries regarding translations into other languages or commercial reproduction or distribution should be addressed to:
NISO, 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211.

ISBN: 978-1-937522-66-7

Contents

Foreword	v
Section 1: Introduction	1
1.1 Purpose and Scope	1
1.2 Terms and Definitions.....	1
Section 2: Recommendations	3
2.1 Package.....	3
2.1.1 Package Content	3
2.1.2 The Manifest Document	3
2.1.3 Package Structure	3
2.1.4 Package Format	5
2.2 Exchanging Packages	5
2.3 Conformance Levels.....	5
2.3.1 Conformance Level 0	5
2.3.2 Conformance Level 1	6
2.3.3 Conformance Level 2	7
2.4 Notes on Conformance Levels	7
Section 3: Guidelines for Effective Implementation	8
3.1 Material Inclusion/Exclusion	8
3.2 Validation	8
3.3 File Naming.....	8
3.4 Consistency	8
3.4.1 Consistency in File Naming	8
3.4.2 Consistency of Delivered File Types	9
3.4.3 Consistency with Serial Item Identifiers.....	9
Appendix A: Examples of Packages	10
Appendix B: Additional Detail on Use Cases	16
Appendix C: How to Apply for ISSN and DOI	20
Appendix D: Related Standards and Recommended Practices	23
Appendix E: FAQ	26
Appendix F: XSD Schemas for each PESC Conformance Level	27

Foreword

About this Recommended Practice

Many different organizations, such as libraries, archives, indexing services, content aggregators, publishers, and content creators exchange and work with the diverse digital files that comprise serial content. There are many reasons for copies of serial content to be transferred from organization to organization, and even within a single organization, many times during the lifecycle of the content. When exchanging content, the files that comprise a serial “publication” are packaged together in some manner and these packages can be highly variable.

There are a number of elements that may be present in the exchange of serial content. These include:

- The content itself, represented by digital files
- Descriptive metadata describing the intellectual content of the files
- Structural metadata describing how the files are related to one another
- A manifest describing what files should be present when more than one file is sent
- The package that holds all of the information described above

These elements can be combined and delivered in many different ways. For example, members of the NISO Protocol for Exchanging Serial Content Working Group have received or delivered such content in the following ways:

- an e-mail with multiple attachments (an XML file along with 3 JPEG files) that represent a single manuscript according to the text of the e-mail;
- a ZIP file that contains hundreds of separate files that may or may not be accompanied by metadata or a manifest file that explains their relationships;
- a tar file that represents an entire issue of content with directory structures that implicitly define relationships between files; and
- an EPUB, comprised of HTML, XML, and image files.

In addition to the variety of delivery methods and packaging, the packages themselves may contain a handful of files for a single issue or over a million files for a full journal title backfile.

If the amount of material being exchanged is small, or the exchange happens rarely, this variation might not be a problem, but for organizations that exchange large quantities of serial content on an ongoing basis, this variation leads to frustrations and inefficiencies.

To address these issues, NISO approved the formation of a Working Group in June 2013 with the charge to:

- Examine current practice and see what strategies are currently used in the community. Depending on the results of this exploration the Working Group might find that it is possible to either adopt a current method wholesale or to make changes to a current method to better address the specific needs of serial content, rather than creating a new method from scratch.
- Develop a Recommended Practice for the specifications of packaging that can be used for archiving and exchanging digital files related to periodic publications.

The guidelines in this document provide the recommendations developed by the PESC Working Group on the best way to manage all of the elements of serial content packaging in a manner that aids both the content provider and the content recipient in understanding what has been delivered and received.

NISO CCM Topic Committee Members

The Content and Collection Management (CCM) Topic Committee had the following members at the time it approved this Recommended Practice:

Marti Heyman, Co-chair
Cengage Learning

Betty Landesman, Co-chair
University of Baltimore

Éva Bolkovac
Yale University Library

Eric Childress
OCLC Research

Marjorie M.K. Hlava
Access Innovations / Data Harmony

Rebecca Kennison
K|N Consultants

Amy Kirchhoff
ITHAKA/JSTOR/Portico

Robert Klingenberg
Yale University Libraries

NISO PESC Working Group Members

The following individuals served on the NISO Protocol for Exchanging Serial Content (PESC) Working Group, which developed and approved this Recommended Practice:

Jean-Baptiste Bertrand
Chargé de système d'information
documentaire - Cléo

Mark Donoghue
IEEE Publishing Technology

Beth Friedman
Data Conversion Laboratory, Inc.

Leslie Johnston, Co-chair
National Archives and Records
Administration

Laurie Kaplan
Serialsolutions/ProQuest

Amy Kirchhoff
ITHAKA/JSTOR/Portico

Andrea Kudzia (through January 2014)
Cengage

Jennifer Melinn (through December 2013)
Wolters Kluwer

Oliver Pesch
EBSCO

Laura Randall (from June 2014)
National Library of Medicine

Gabriele Schaefer (from February 2014)
Springer

Henning Schönenberger (through January 2014)
Springer

David Schott (through February 2014)
Copyright Clearance Center, Inc.

Linda Sussman
Cold Spring Harbor Laboratory Press

Lauren Syer (through February 2014)
Harvard University

Kimberly A. Tryka, Co-chair
National Institute of Standards and
Technology

Stephen Want
U.S. Copyright Office

Elizabeth Windsor
Project MUSE

Amy Wood
Center for Research Libraries

Wei Zhao
OCUL/Scholars Portal

Acknowledgements

The PESC Working Group wishes to acknowledge those outside the formal working group membership who contributed to this effort:

Marti Heyman (Cengage Learning), Nettie Lagace (NISO), and Cynthia Hodgson (NISO)

Trademarks, Service Marks

Wherever used in this standard, all terms that are trademarks or service marks are and remain the property of their respective owners.

Section 1: Introduction

1.1 Purpose and Scope

Serial publications represent a diverse content space ranging from popular magazines to scholarly journals, from content that is image-based to content that is text-based, from publications that have new content daily (even hourly) to those that might have new content only every few years. The manner in which the content creator packages the content for any particular serial publication may not match the needs of the recipient of that content and, given the proliferation of serial content, this creates a highly chaotic environment with different partners delivering significantly different looking content packages and making it difficult for all parties in the environment to work with content deliveries in a cost-efficient and effective way.

The recommendations in this document offer guidance to members of the scholarly communication community on preferred practices for the packaging and exchange of serial content that will enable the automation of processes to receive and manage serial content at scale. By following these practices, organizations can make it clear what content has been transmitted, how it is organized, and what processing is required when a new package is received.

1.2 Terms and Definitions

The following terms, as used in this recommended practice, have the meanings indicated.

<u>Term</u>	<u>Definition</u>
actors	Sending and receiving parties (see Appendix B.2).
BagIt	A hierarchical file packaging format designed to support disk-based or network-based storage and transfer of arbitrary digital content. <i>Source:</i> BagIt specification at: http://tools.ietf.org/html/draft-kunze-bagit-10
checksum	A short piece of text that is algorithmically derived from a file and allows the determination of whether the file has been corrupted.
conformance	Compliance with the guidelines of this recommended practice. <i>Note:</i> Conformance is defined at different levels.
descriptive metadata	Metadata that describes the intellectual content of a file, as opposed to the physical or structural characteristics of the file.
digital object identifier DOI®	A unique, persistent, and actionable digital identifier for a content object. <i>Note:</i> The DOI syntax is defined in ANSI/NISO Z39.84 and the entire DOI system is defined in ISO 26324.
full-text XML	Bibliographic metadata plus the full text of the article is included in an XML file.
ISSN	International Standard Serial Number, an eight digit number assigned to a continuing resource, according to the standard ISO 3297.

<u>Term</u>	<u>Definition</u>
manifest	Detailed listing of the items contained in a package.
metadata	A set of data that describes and gives information about other data; for example, data describing the full text of an article.
package	An aggregation of files being sent to another actor. <i>Note:</i> In the Open Archival Information System (OAIS), this package is often the Submission Information Package (SIP).
rendition	Electronic encoding of content. <i>Note:</i> Any individual article may have one or more renditions. Sometimes these are provided by the content provider, though sometimes they must be created by the recipient. For example, the following are all different renditions of the same content: • A PDF file of page images of the article • A set of TIFF images of each page of the article • A content owner-provided HTML file of the article • An HTML file that is created on the fly from full-text XML provided by the content owner
serial	A continuing resource issued in a succession of discrete issues or parts, usually bearing numbering, that has no predetermined conclusion. <i>Examples:</i> Journals, magazines, electronic journals, ongoing directories, annual reports, newspapers, monographic series, and also those journals, magazines, and newsletters of limited duration that otherwise bear all the characteristics of serials (e.g., newsletter of an event). <i>Source:</i> ISO 3297
serial item	The smallest self-contained portion of a serial. <i>Examples:</i> Articles, letters, reviews, case studies, front matter (in contrast to volumes or issues)
structural metadata	Metadata that is used to illustrate the relationships between a series of files and that allows the reconstruction of a whole from its parts.
tar	A file format that groups many files together while retaining their relationships within the file system.
use cases	Scenarios in which content is exchanged between actors.
ZIP	A file format that allows many files to be grouped together and compressed.

Section 2: Recommendations

2.1 Package

The package includes all of the content being sent along with other information needed to understand those files, as described in [2.1.1](#) through [2.1.4](#). Minimally, it must include content files and a manifest file. It may also include descriptive and structural metadata.

2.1.1 Package Content

The package may contain three types of information:

- Manifest – The manifest document describes the package and its content. It is required for all Conformance Levels (see [2.3](#)).
- Content Files – Examples of content files are: full text of a serial item (PDF or XML), images, supplementary materials, metadata, and cited references.
- Metadata Files – Metadata files describe the content items being delivered. This is required only at Conformance Level 2 (see [2.3.3](#)).

2.1.2 The Manifest Document

A manifest document must be contained in all packages and should be named “manifest” with the appropriate file extension for the format used. The manifest will include information related to the package as a whole and will list the files included (as well as their directory structure if there is one), with information about the files corresponding to the Conformance Level of the package (see [2.3](#)).

The encoding of the manifest.txt file should be clearly stated. If using the BagIt packaging this would be stated in the bagit.txt file. If using XML the encoding is assumed to be UTF-8 unless explicitly stated otherwise in the XML declaration.

2.1.3 Package Structure

It is helpful for package recipients to have sufficient information in the package structure and file naming conventions to identify the set of serial items being delivered. To this end, the files within the package should be organized into a folder structure (see [Figure 1](#)) that represents the logical organization of the content as follows:

- Every package should have a manifest file that appears in the root directory of the package. It should be named “manifest” and have the proper extension for its format. (For example, manifest.txt or manifest.xml.)
- If multiple serial items are in the package, the contents of each item shall appear in a separate folder. To ensure unique folder names within a package, the folder names can be derived from the item’s DOI® (or other identifier).
- If multiple sets of serial items (for example, journal issues) are included in a single package, each set shall be represented by a separate folder containing the serial item folders. In general, it is advisable for the hierarchy of the serial publication to be represented in the folder structure. (For example, an issue folder might be named with the journal ISSN, publication year, volume number, and issue number, and for an article in a journal without issues or volume, then the directory structure and file naming would contain the ISSN and

publication year). If multiple issues are included in the package, each issue shall be represented by a separate folder and should represent the volume and issue of the material being sent. When “ahead of print” materials are exchanged for serials that use volume/issue enumeration, it is recommended to group the “ahead of print” materials using volume/issue designation of 0.

- If multiple journals are included in the package, each journal shall be represented by a separate folder with that journal’s issues and articles appearing in sub-folders. The folder name for a journal should be the ISSN or other unique identifier of that journal as used in the package.

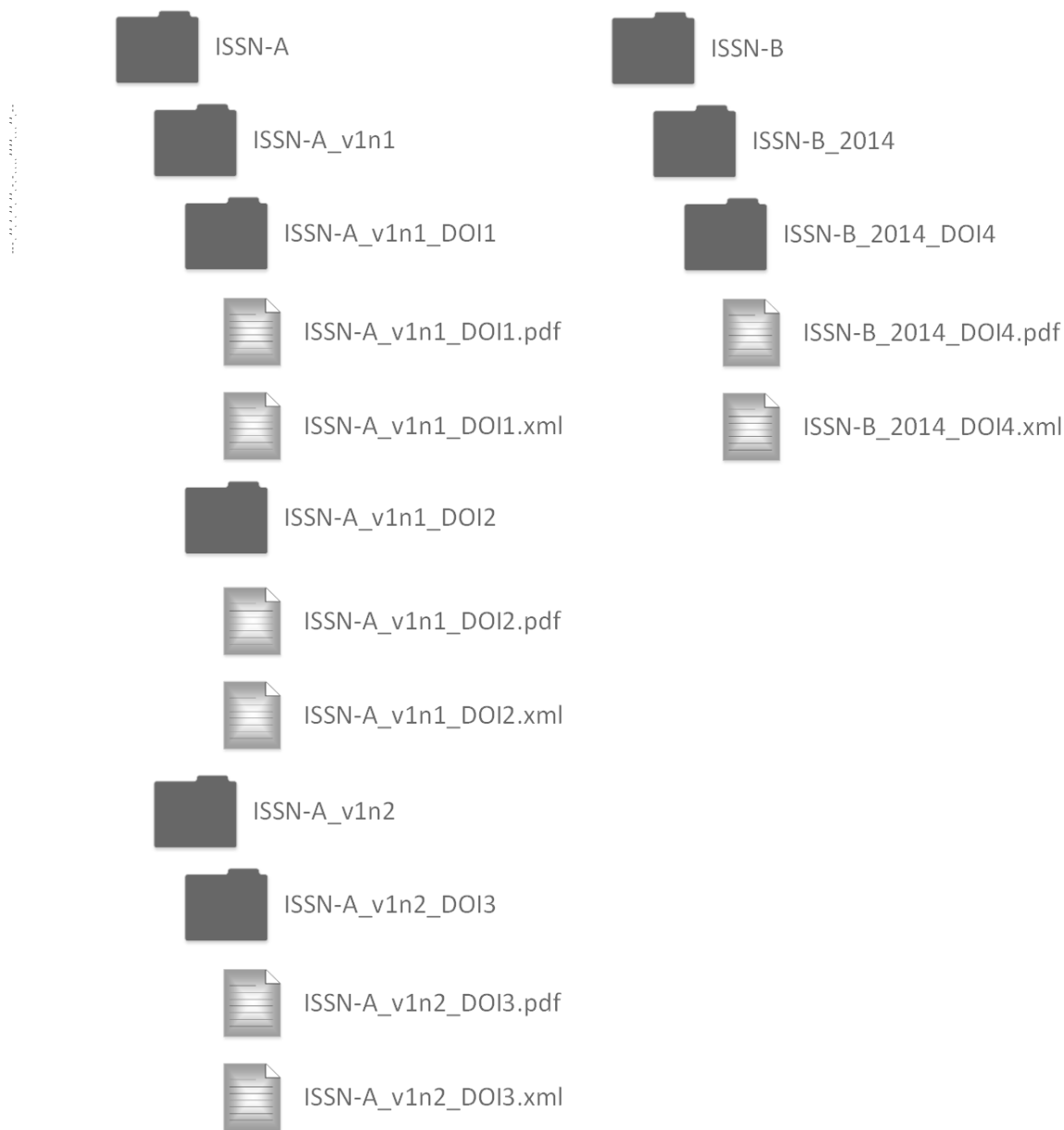


Figure 1: Example schematic of package structure

2.1.4 Package Format

The content being exchanged should be packaged using ZIP, tar, BagIt, or other commonly used formats. A suitable format will:

- Allow multiple files to be grouped within a single container (or file, for example a ZIP file) that will be used in the exchange
- Support a folder or directory structure so files can be logically grouped within the package
- Be mutually agreeable to both parties involved in the exchange

In the experiences of the PESC Working Group members, a package that contains no more than 100 articles is advisable.

2.2 Exchanging Packages

Organizations that receive large numbers of files from large numbers of partners over a prolonged period of time are likely to set up FTP (or SFTP) accounts for their partners to transmit the files. Other common transmission methods are rsync or OAI-ORE. These particular methods ensure that packages are sent to a pre-determined destination that can be automatically monitored for the arrival of new content or, in the case of OAI-ORE, packages can be fetched.

In the case of small packages sent on a one-off basis, it might be possible that both partners would consent to e-mail exchange rather than bother with the creation of FTP accounts. In general, this is not advisable as it is very easy for content to get lost in e-mail accounts and traceability is very difficult.

In rare instances where a partner might be sending very large amounts of content, such that the use of the regular FTP protocol would result in days of transfer, it might become practical to look into either high-speed file transfer software (such as Aspera) or shipping hard drives. This would need to be negotiated between the two organizations, including the selection of hard drives.

2.3 Conformance Levels

Level 1 should be considered the default level of conformance. Level 0 will allow adequate exchange between organizations and could be the best choice if one, or both, partners in the exchange don't have a workflow capable of easily producing or consuming the information in Level 1. Level 2 is only likely to be used in cases where both parties agree that a greater depth of information exchange will be of mutual benefit.

Examples of all Conformance Levels, and the package file structure they correspond to, are shown in [Appendix A](#).

2.3.1 Conformance Level 0

For a package to conform to Level 0 it must contain the content files being sent as well as the following information in a distinct manifest document.

For the entire package:

- The Conformance Level of the package (0, 1, or 2)
- Creation date of the package

- The default purpose of the package (suggested values are new, replace, version, delete) – This can be superseded for individual items.
- Sender name and contact information
- What serial items are included

For each serial item:

- The files that belong to it

For each file:

- Its relative location in the package, i.e., the directory or folder structure
- The purpose of the item, if it differs from the default purpose of the package
- A unique identifier – A DOI or other globally unique identifier is preferred
If a DOI is not available, use a proprietary ID that uniquely identifies work or item. The unique identifier facilitates effective processing of replace, version, and delete actions on items in subsequent packages. If using an identifier “type” tag (see examples in [Appendix A](#)), it is recommended to use the list of types from the Journal Article and Interchange Tag Library for the <pub-id-type> attribute (see <http://jats.nlm.nih.gov/archiving/tag-library/1.1d1/?attr=pub-id-type>).

2.3.2 Conformance Level 1

Conformance to Level 1 requires all the information required for Level 0 (content files and manifest file) with the additional information as noted below added to the manifest file.

For each individual file:

- Media type (in accordance with RFC 6838, *Media Type Specifications and Registration Procedures*)
- Checksum and checksum type (for example, an MD5 or SHA-1 cryptographic hash)
- Relationship to other files in the package. This is indicated by a <role> element in the manifest. It is expected that parties involved in an exchange of data will collaborate on defining appropriate roles. A list of suggested roles, as used in the sample files, is given below:
 - component: figure graphic
 - component: formula graphic
 - component: marked up header
 - component: media
 - component: other
 - component: other graphic
 - component: page images
 - component: supplemental file
 - component: table graphic
 - data: marked up