



Version 1.0

**The PRISM Source Vocabulary
Specification Overview**

September 27, 2012



Copyright and Legal Notices

© 2001 – 2012 International Digital Enterprise Alliance, Inc. All Rights Reserved.

nextPub® is a registered trademark of the International Digital Enterprise Alliance, Inc. (IDEAlliance).

This document may be downloaded and copied provided that the above copyright notice and this Notice are included on all such copies. This document itself may not be modified in any way, except as needed for the purpose of developing International Digital Enterprise Alliance, Inc. ("IDEAlliance") specifications. Use of the specification or standard set forth in this document shall not create for the user any rights in or to such specification or standard or this document, which rights are exclusively reserved to IDEAlliance or its licensors or contributors.

Use of this document and any specification or standard contained herein is voluntary. By making use of this document or any specification or standard contained herein, the user assumes all risks and waives all claims against IDEAlliance, its licensors and contributors. By making this document available, IDEAlliance is not providing any professional services or advice to any person or entity. Any person or entity utilizing this document or any specification or standard contained herein should rely upon the advice of a competent professional before using any such information.

NO WARRANTY, EXPRESSED OR IMPLIED, IS MADE REGARDING THE ACCURACY, ADEQUACY, COMPLETENESS, LEGALITY, RELIABILITY OR USEFULNESS OF ANY INFORMATION CONTAINED IN THIS DOCUMENT OR IN ANY SPECIFICATION OR STANDARD OR OTHER PRODUCT MADE AVAILABLE BY IDEALLIANCE. THIS DOCUMENT AND THE INFORMATION CONTAINED HEREIN AND INCLUDED IN ANY SPECIFICATION OR STANDARD OR OTHER PRODUCT OR SERVICE OF IDEALLIANCE IS PROVIDED ON AN "AS IS" BASIS. IDEALLIANCE DISCLAIMS ALL WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, ANY ACTUAL OR ASSERTED WARRANTY OF NON-INFRINGEMENT OF PROPRIETARY RIGHTS, MERCHANTABILITY, OR FITNESS FOR A PARTICULAR PURPOSE.

IN NO EVENT SHALL IDEALLIANCE, ITS LICENSEES, CONTRIBUTORS OR THEIR RESPECTIVE OFFICERS, DIRECTORS, EMPLOYEES, AGENTS, REPRESENTATIVES, SUPPLIERS OR CONTENT OR SERVICE PROVIDERS BE LIABLE FOR DAMAGES OF ANY KIND, INCLUDING WITHOUT LIMITATION, DIRECT, INDIRECT, COMPENSATORY, SPECIAL, INCIDENTAL, PUNITIVE OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FROM DATA LOSS OR BUSINESS INTERRUPTION) EVEN IF MADE AWARE OF THE POSSIBILITY OF SUCH DAMAGES, WHETHER IN AN ACTION UNDER CONTRACT, TORT OR ANY OTHER THEORY, ARISING OUT OF OR IN CONNECTION WITH THE USE, INABILITY TO USE OR PERFORMANCE OF THIS DOCUMENT, THE SPECIFICATION OR STANDARD CONTAINED HEREIN, OR ANY OTHER DOCUMENT OR SPECIFICATION OR STANDARD MADE AVAILABLE BY IDEALLIANCE.

Some states do not allow the disclaimer or limitation of damages, so the disclaimers set forth above apply to the maximum extent permitted under applicable law.

IDEAlliance takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed or implicated with respect to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available. IDEAlliance does not represent that it has made any effort to identify any such rights. Information on IDEAlliance's procedures with respect to rights in IDEAlliance specifications can be found at the IDEAlliance website at www.idealliance.org. Copies of third-party claims of rights, assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification, can be obtained from the President of IDEAlliance at patent-disclosure@idealliance.org.

IDEAlliance requests interested parties to disclose any copyrights, trademarks, service marks, patents, patent applications, or other proprietary or intellectual property rights which may cover technology that may be required to implement this specification. Please address the information to the President of IDEAlliance at patent-disclosure@idealliance.org.

Abstract

On April 3, 2010 the publishing ecosystem was changed forever. For magazines, the launch of the iPad signaled the exciting new era of interactive, innovative publications and almost unimagined new revenue opportunities. But forced to use today's workflows, tools and technologies, producing multiple tablet editions translates into escalating production staff efforts without any accommodation to publication schedules. The existing print-based workflow was unsustainable and unprofitable, so the urgency to find solutions could not have been greater!



nextPub[®], the publishing industry's technology incubator, is an industry initiative hosted by IDEAlliance and supported by MPA, the Magazine Media Association, AAAA/Ad-ID and the Japanese Magazine Publishing Association. By enabling publishers to engage directly with their industry peers, nextPub brings together the industry's best minds and provides the additional technical and support resources required to aggressively tackle the challenge. The goal of nextPub is to foster the development of next-generation publishing tools by embracing emerging technologies, developing best practices and establishing new industry specifications to make multi-channel publishing both simple and efficient.

Those engaged in the nextPub Initiative believe the solution to efficient multi-channel publishing is in the *SOURCE*. nextPub is therefore publishing a set of foundational content management specifications, built on PRISM 3.0 metadata elements and employing HTML5 for semantic content encoding.

Sponsors Who Support nextPub Development



 Ad-ID Advertising Identification and Management	 Adobe	 JMPA	 MEDIA BEACON®
 AGFA	 Time Inc.	 MarkLogic	 REALLY [] STRATEGIES™ Bringing Strategy, Content & Technology Together
 APEX	 mpa	 Typéfi	 zinio™
 Aysling	 DALIM SOFTWARE	 PROIMAGE Multi Channel Publishing Solutions	 WOODWING
 HEARST Magazines	 HIPZONE Magazine Distribution	 mirasol Displays by Qualcomm	

Table of Contents

1	STATUS.....	1
1.1	DOCUMENT STATUS	1
1.2	DOCUMENT LOCATION	1
1.3	VERSION HISTORY	1
2	INTRODUCTION	2
2.1	THE TWO WORLDS OF PUBLISHING	2
2.1.1	The World of Content-Based Publishing.....	2
2.1.2	The World of Design-Based Publishing	2
2.2	THE NEXTPUB CHALLENGE	3
2.3	DESIGNING A DYNAMIC CONTENT ARCHITECTURE	3
2.4	THE NEXTPUB PUBLISHING MODEL	2
3	ABOUT THE PRISM SOURCE VOCABULARY	3
3.1	PSV – A FLEXIBLE FRAMEWORK.....	3
3.2	IMPLEMENTATION OPTIONS.....	4
3.2.1	Scenario #1; A Simple Magazine Publishing Solution.....	4
3.2.2	Scenario #2: Simple Book/eBook Publishing.....	5
3.2.3	Scenario #3: Tracking Usage and Usage Rights for Magazines	6
3.3	REQUIRED PSV BUILDING BLOCKS.....	7
3.4	MORE ABOUT PSV METADATA BUILDING BLOCKS.....	7
3.4.1	Description.....	7
3.4.1.1	About Subjects.....	8
3.4.1.2	About Genre.....	8
3.4.2	Where Used	8
3.4.3	Usage Rights.....	9
3.4.4	Relations	10
3.4.5	Components.....	11
3.5	PSV CONTENT MANAGEMENT SCHEMA.....	11
3.6	OTHER PSV SCHEMAS.....	11
3.7	ABOUT ADVERTISING.....	12
3.8	PSV CONTENT BUILDING BLOCKS.....	12
3.8.1	Benefits of Employing HTML5.....	12
3.8.2	Optional HTML5 Head Element.....	13

3.8.3	Required HTML5 Body Element	13
3.8.3.1	Article as the Root Element	14
3.8.3.2	Additional Semantic Tags	14
3.8.3.3	Figure and Rich Media Tags	14
3.8.3.4	Adding Semantics Using the HTML5 Class Attribute.....	15
4	RELATIONSHIP TO OTHER SPECIFICATIONS	16
4.1	RELATIONSHIP TO PRISM	16
4.2	RELATIONSHIP TO PAM	16
4.3	RELATIONSHIP TO HTML5	17
4.4	RELATIONSHIP TO EPUB 3.0	17
5	PSV NAMESPACE, DOCUMENTATION AND CM SCHEMA.....	18
5.1	THE PSV NAMESPACE.....	18
5.2	THE PSV VERSION 1.0 DOCUMENTATION PACKAGE	18
5.3	THE PRISM DOCUMENTATION PACKAGE	19
5.3.1	General Documents.....	19
5.3.2	PRISM Metadata Specifications	19
5.3.3	PRISM Controlled Vocabulary Specifications	20
APPENDIX A	NORMATIVE REFERENCES	21
A.1	NORMATIVE REFERENCES FOR PRISM AND PSV SPECIFICATIONS.....	21
A.2	NON-NORMATIVE REFERENCES FOR PRISM AND PSV SPECIFICATIONS.....	23

1 STATUS

1.1 Document Status

This document represents the IDEAlliance PRISM Source Vocabulary Overview Specification

✓	Draft for Public Comment	December 12, 2011
✓	V1.0 PSV Final Draft Specifications	June 12, 2012
✓	V1.0 PSV Final Specifications	September 27, 2012

1.2 Document Location

The location of this document is:

http://www.prismstandard.org/specifications/psv/1.0/PSV_overview.pdf

or

http://www.prismstandard.org/specifications/psv/1.0/PSV_overview.htm

1.3 Version History

Version Number	Release Date	Editor	Description
3.0 Public Draft	12/15/2011	Kennedy	Version to support nextPub
3.0 Final Draft	06/12/2012	Kennedy	Comments Resolved, Final Draft
3.0 Final Draft	09/27/2012	Kennedy	Comments Resolved, Final V1.0 Specification

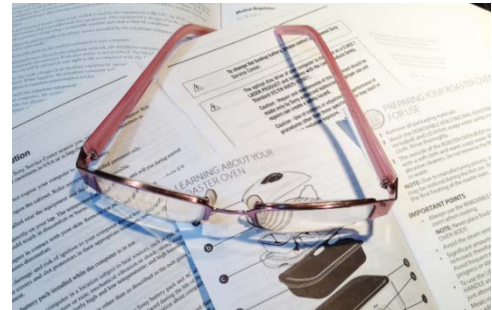
2 INTRODUCTION

2.1 The Two Worlds of Publishing

When publishing moved from linotype to computer assisted composition, the publishing world was split into two different and distinct worlds. These worlds adopted different tool sets, different workflows and clearly different philosophies.

2.1.1 The World of Content-Based Publishing

The world of content-based publishing is commonly known as technical publishing. In this world, the content is the whole point of publishing. Technical publishing encompasses technical documentation such as maintenance and operational technical manuals as well as reference publishing such as legal publishing, financial reports and product information along with content that is assembled from databases such as course catalogs, racing forms and even television show guides. Content-based publications are often quite lengthy, being made up of hundreds and often thousands of pages of highly structured content.



Another characteristic of this type of publication is that it is typically formatted automatically by applying style sheets or templates. Style sheets are designed to enable a reader to quickly access information in lengthy publications.

Publishers with large volumes of structured content adopted publishing tools and systems based on ISO 8879:SGML and later on W3C's XML markup language. Content could be created in XML first, stored in XML content management systems that were modeled on the XML document structure. Content could then be assembled from the XML content repositories and the layout and styling were automated through the use of high-speed computerized pagination engines.

2.1.2 The World of Design-Based Publishing

A second and very different world of publishing is based on design. Most magazines fall into this category along with highly designed books such as cook books, travel books and other “coffee table” books. For this type of publishing, design comes first. The Art Director often develops the concept and design before any content is created. Standard style sheets that give each issue the same look month after month would simply not do. And because design comes first, content structures cannot be effectively standardized either. Hence, very few publishers of design-based publications have



adopted “XML-first” workflows. And in fact until recently very few members of this community employed XML at all. Process automation has been held to the minimum as well.

2.2 The nextPub Challenge

Until the tablet-publishing tsunami hit, design-based publications were able to justify their labor-intensive design-based publication process. But the challenge of producing well-designed publications, not only in print, but on a growing number of digital devices with different resolutions, aspect ratios and sizes soon became overwhelming. The challenge to the nextPub participants was to figure out how to leverage the benefits realized from the XML-First publishing model while protecting the seminal importance of high-quality design.

2.3 Designing a Dynamic Content Architecture

When the nextPub Initiative began, the immediate goal was to make multi-channel publishing both simple and efficient. That is, to meet the iPad challenge. But as the Working Group explored standards and technologies, strategists began to envision monetizing content beyond today’s publishing channels. Strategists began talking about developing a “Dynamic Content Architecture” from which new publishing channels can spring by assembling and delivering new content collections based on topics pulled from a multitude of magazine titles or based on personal preferences foretold by Flipboard, Pulse or Zite models.

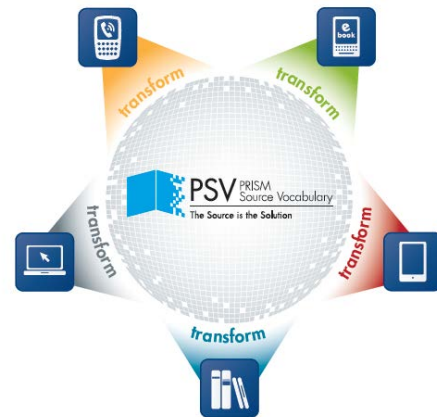
Based on this more aggressive vision, nextPub developed a set of design principles:

1. We have come to believe the *Source is the Solution*. We must capture and store platform-agnostic content as early as possible. For design-based publications, this will likely *not* mean *XML First*, but it certainly implies *XML Early*.
2. Since design-based content *cannot* be highly structured, we need to focus on using metadata to enable the management and assembly of content objects. This means that the XML will be flexible and semantically based. This means that content repositories cannot be based on document structures, but rather must be based on semantic metadata.
3. Since design-based content is not highly structured, the framework for a dynamic content architecture must be modular and flexible.
4. We cannot afford to throw away existing systems and standards that we already have in place; so we must build on them.
5. We must plan for the future embracing emerging technologies.
6. We must foster the development of new dynamic display/layout technologies that enable simultaneous design for multiple publishing platforms and automatic content layout capabilities supporting high-quality design aesthetics.

2.4 The nextPub Publishing Model

The nextPub Publishing Model is based on digital capture and management of all content and associated rich media. Source content must be semantically rich enough to enable the publisher to select content and automate layout and delivery to a wide variety of publishing platform in platform-native formats. To support the nextPub publishing model, the Working Group has developed the PRISM Source Vocabulary (PSV) Specification. PSV has been designed to support today's issue-based publications as well as to enable publishers to aggregate content in new ways to create new digital content channels that extend beyond the traditional publication models.

The PRISM Source Vocabulary (PSV) Specification, defines a framework of robust metadata elements and controlled vocabularies that can be used to configure federated source content / rich media repositories. Metadata fields and values used in this specification are drawn from the IDEAlliance PRISM 3.0 Metadata and Controlled Vocabulary Specifications. In order to future-proof content, nextPub recommends encoding content using HTML5 tagging enhanced with PRISM-based class semantics.



3 ABOUT THE PRISM SOURCE VOCABULARY

When designing the PRISM Source Vocabulary, the nextPub Working Group followed the basic principles outlined in Section 2. We are committed to build on existing publishing infrastructures. This meant basing our work on the PRISM (Publishing Requirements for Industry Standard Metadata) Specification which is already in wide use by design-based publishers. Since most major magazine publishers have implemented today's publishing systems based on this robust IDEX metadata specification, it was deemed to be the ideal starting point to build on for the future.

We are also committed to leveraging emerging technologies. Most of the current publishing systems based on PRISM were using an XHTML-based content coding scheme that publishers were using to deliver content to aggregators. But the nextPub Working Group found great limitations in remaining with XHTML and great opportunity to move to HTML5, a semantically-based "Vocabulary" for HTML and XML [HTML5]. HTML5, working in conjunction with CSS3 and JS, provided the support for rich media and interactivity that design-based publishers require for the future.



Figure 3.1 PSV Model

Note: While some from the XML-First, content-based publishing world may question the use of HTML5 based on what they view as a limited functionality due to the high degree of structural flexibility and limited hierarchy of HTML5, it is that very flexibility plus the ability to carry semantics on any structure tag that makes HTML5 a perfect fit for design-based content.

3.1 PSV – A Flexible Framework

Another principle behind the design of PSV was the requirement for flexibility. Each publisher has their own business models and publishing strategies. So a single, rigidly defined framework will not work for anyone. Rather, the PSV design provides a framework and the building blocks required to implement a dynamic publishing system that can easily be tailored for the business requirements of each publisher. By selecting appropriate PSV metadata building blocks, a publisher can build a federated digital content repository that satisfies their unique business requirements and strategies. And the publisher is free to use any conformant HTML5 to semantically encode source content.

Note: A minimum set of PSV building blocks are required. All others are optional and will be selected based on the publishers' business requirements. See Figure 3.2.

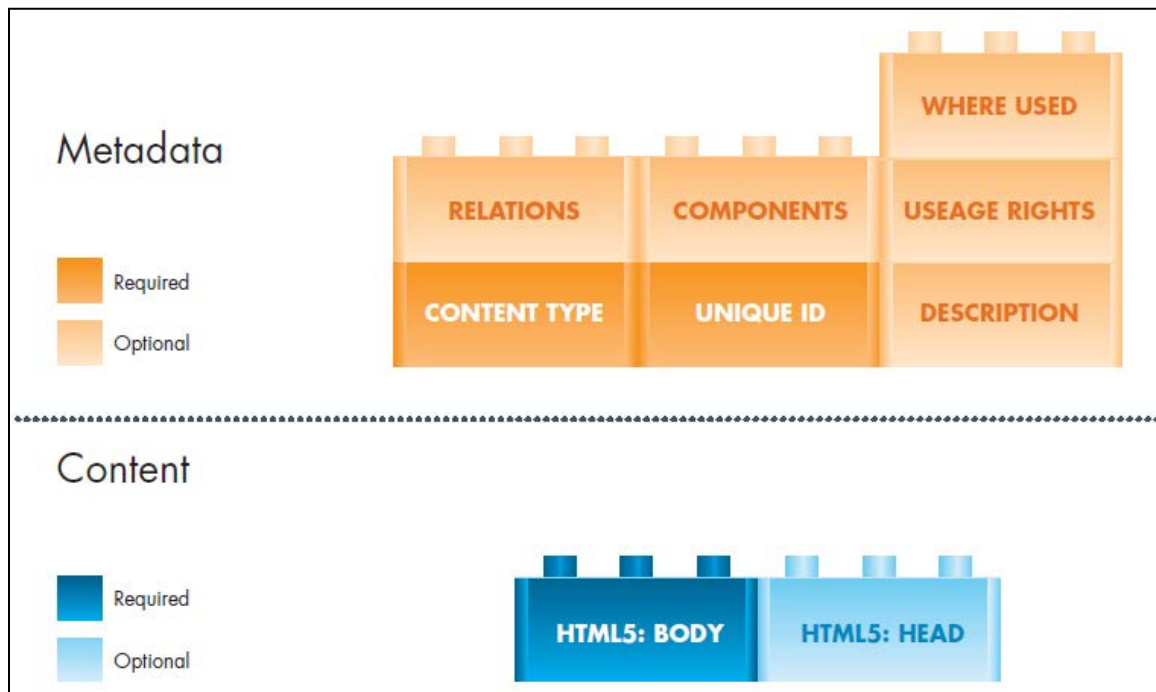


Figure 2.1 PSV Framework Building Blocks

3.2 Implementation Options

PSV's modular design provides the building blocks for a publisher to implement a dynamic, cross-platform publishing system. PSV can easily be tailored for the business requirements of each publisher.

3.2.1 Scenario #1; A Simple Magazine Publishing Solution

In the first scenario a publisher wants to move from publishing a magazine in print to publishing both in print and on a tablet. The publisher wants to implement a simple standards-based content repository from which content can be assembled into a print and into a digital edition on a monthly basis. The publisher would like to add complexity and functionality to the publishing system over time, but prefers to start "simple."

This simple publishing solution can be built by using just the PSV required metadata blocks and simple HTML5 body tags to encode content and rich media. In order to provide a mechanism for the publisher to find content in the PSV repository, metadata fields such as title, subject, keywords, etc. from the descriptive metadata block will be used. In addition some fields from the "Where Used" block must be included so that the publisher can encode the publication, issue and article identification fields. See Figure 3.1.

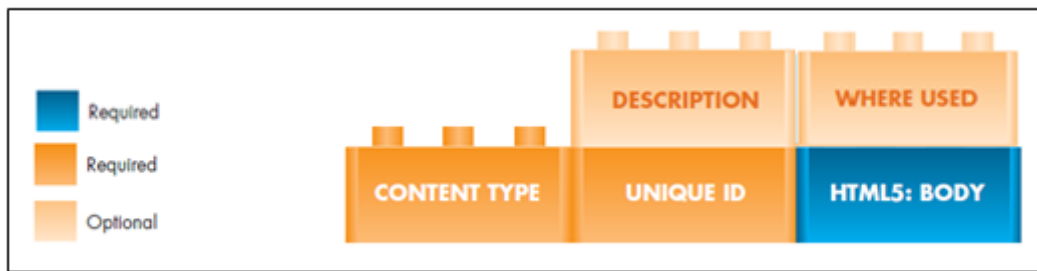


Figure 3.1 Scenario #1: A simple PSV Implementation for magazine publishing

For this simple implementation of PSV, we would create a schema that includes only those building blocks and fields required by this publisher as a starting point for magazine publishing. In addition to selecting just the building blocks the publisher requires, we have also deleted metadata fields from PSV building blocks that this publisher does not wish to capture or track. For example a magazine publisher would likely not track an ISBN, a copyright year or a national catalogue number. See Figure 3.2.

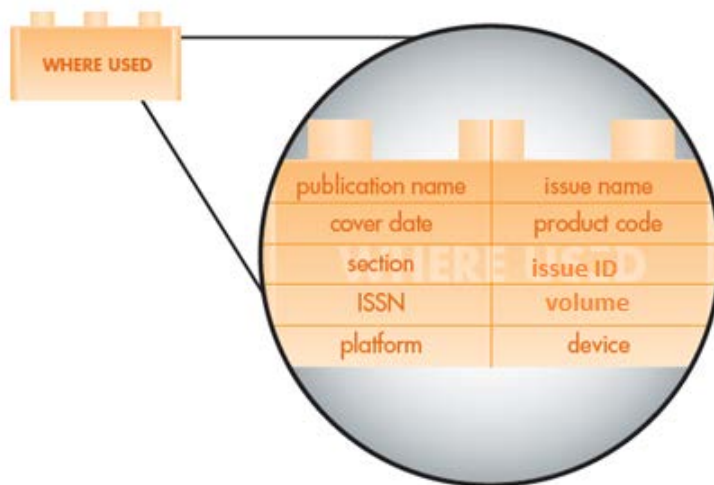


Figure 3.2 PSV Metadata for a magazine

3.2.2 Scenario #2: Simple Book/eBook Publishing

In this scenario a publisher wants to move from publishing books in print to publishing both in print and as an EPUB3 eBook. The publisher wants to implement a simple standards-based content repository from which content can be assembled into a print and into EPUB3 eBooks. The publisher would like to add complexity and functionality to the publishing system over time, but prefers to start “simple.” See Figure 3.3.

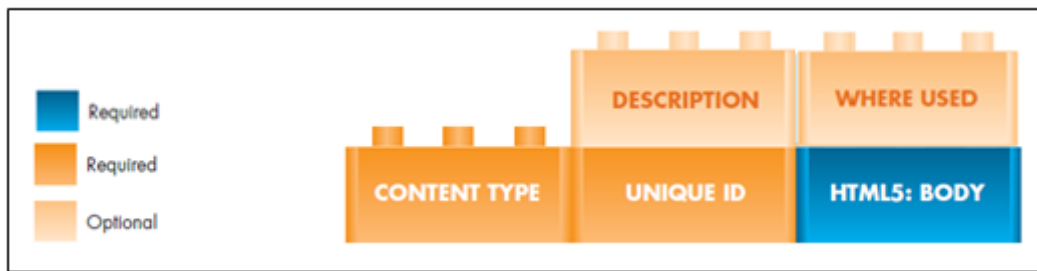


Figure 3.3 Simple Book Publishing Scenario

For this simple implementation of PSV, we would create a schema that includes only those building blocks and fields required by this publisher as a starting point for publishing books in print and as EPUB3 eBooks. These fields will differ from the fields that a magazine publisher would use. For example a book publisher would include book and chapter metadata in the “Where Used” block where the magazine publisher would use the issue and article metadata. See Figure 3.4.

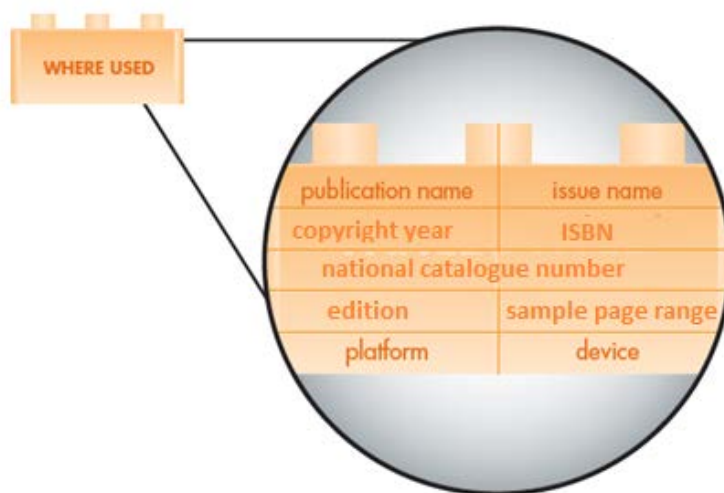


Figure 3.4 PSV Metadata for a Book

3.2.3 Scenario #3: Tracking Usage and Usage Rights for Magazines

In the third scenario, the publisher needs to implement a content repository from which content for several magazine titles can be assembled for both print and the leading tablets. In addition the publisher would like the ability to reuse content from previously published magazines to produce a “Best of 2012” special edition for the year-end. This requirement means that the system must track content usage. Additionally due to concern over usage rights for content across platforms and publications, the publisher requires usage rights tracking as well.

Just as in Scenario #1, the publishing solution begins with the required PSV metadata blocks. In order to provide a mechanism for the publisher to find content in the PSV repository, metadata

fields from the descriptive metadata block must be used. Finally the system must be implemented so that it tracks both usage (where used) and usage rights. See Figure 3.5.

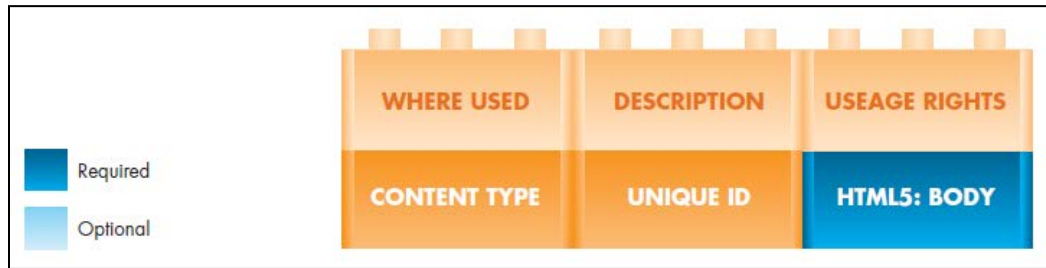


Figure 2.5 A PSV Implementation to track Usage and Rights for Magazines

3.3 Required PSV Building Blocks

In order to manage publishing content in a repository, three elements are required. The first is a unique identifier for each piece of content. You can think of this as a unique name or label by which we can file the content in a repository and that we can use to locate and use the content.

The second element required by PSV is a field to identify the type of content unit you are storing in the repository. The content type field tells us whether we are storing a cover, a table of contents, a masthead, an article or even an advertisement. A PSV Content Type is the highest-level indicator of the nature of the asset that is stored in the repository.

The third required PSV building block is the content itself. PSV uses the body tags from HTML5 for this purpose.

3.4 More About PSV Metadata Building Blocks

In addition to providing the publisher a choice of the metadata blocks with which to implement their system, PSV also provides flexibility within each metadata block. Each major metadata block is made up of component metadata fields from which the publisher may choose.

3.4.1 Description

The fields within the description block are key to designing a system where content is “findable.” Unless you memorize every unique identifier, you need help to find the content you want. The description metadata block offers 30 optional metadata fields to describe content in the PSV repository. Each publisher will select those fields by which they wish to organize content. See Figure 2.6 for some examples of descriptive metadata.

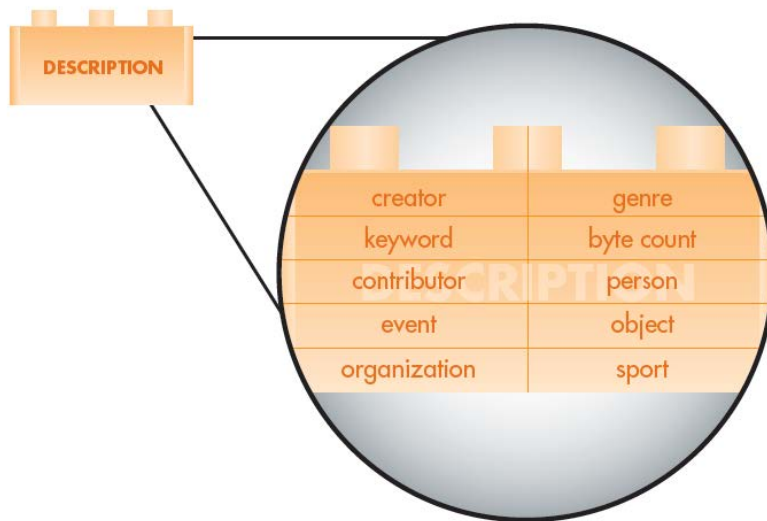


Figure 2.6 Example Description Metadata Fields

3.4.1.1 About Subjects

The subject of content is a key descriptive field. PSV provides a generic “subject” field where the publisher can fill in any text string. Sometimes the generic subject is simply too general to be of much use. PSV provides a number of specialized subject fields to provide for more precise storage and retrieval. These specialized subject fields include academic field, event, person, object, organization, profession, sport and industry.

3.4.1.2 About Genre

Another key descriptive field is the “genre.” Genre is used to refine the definition of the content type of any content or media asset. For example, the following values for genre can be used to refine the content type “article.”

- analysis
- column
- cover story
- department
- essay
- fashion shoot
- feature
- fiction

3.4.2 Where Used

The “Where Used” metadata block allows for the tracking of where content was used. Key choices for the publisher include tracking content usage by the publication name, its ISBN, ISSN,

product code and issue name. PSV allows for tracking the platform and even the device where the content was published. PSV also allows for tracking the section or page of the publication where the content appeared. Altogether, PSV offers nearly 40 optional fields to describe where content was used. Building block for Where Used include publicationInfo, issueInfo, articleInfo, websiteInfo, bookInfo, chapterInfo, blogInfo and blogEntryInfo so PSV can be used for virtually any type of publication. See Figure 2.7.

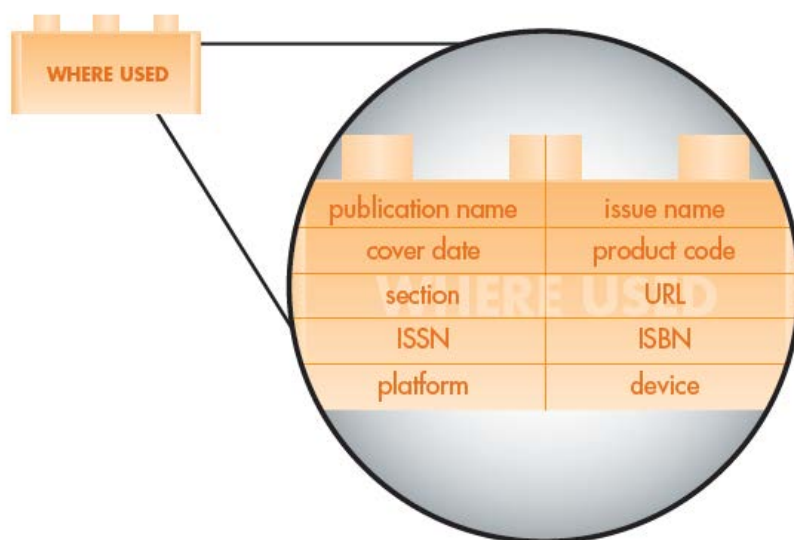


Figure 2.7 Example “Where Used” Metadata Fields

Note that this model allows for tracking multiple uses of the content over time. So for example a recipe may first be published in a magazine article and later in a cook book. The aggregation type field should be used to indicate the type of publication where the content was published.

Aggregation types for PSV content include blog, book, bookazine, catalog, feed, journal, magazine, manual, newsletter, newspaper, other, report, pamphlet, vook and whitepaper.

3.4.3 Usage Rights

When publishing content across platforms and publication titles, the publisher’s right to use the content is not always guaranteed. The Usage Rights metadata block provides optional metadata fields that can be used by publishers to track usage rights of content in a repository. The 15 optional metadata fields in this block are based on the PRISM Usage Rights Metadata Specification –[PRISMURMS]. See Figure 2.8.

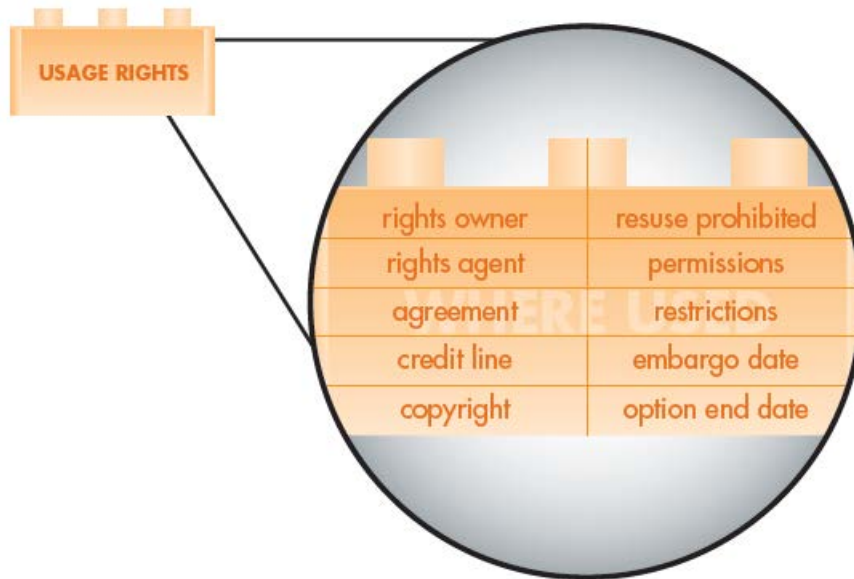


Figure 2.8 Example Usage Rights Metadata Fields

3.4.4 Relations

Sometimes a unit of content is related to other units of content. The Relations metadata block was developed to enable those who wish to track content relationships to do so. The four fields in this metadata block indicate relationships and the nature of the relationships. This field will likely be used when designing highly complex publishing systems. See Figure 2.9.

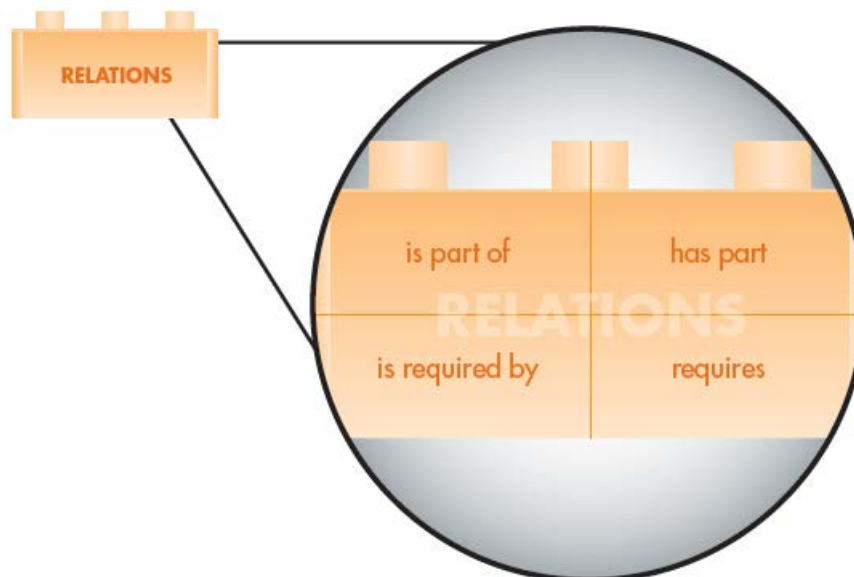


Figure 2.9 Relations Metadata Fields

3.4.5 Components

All the metadata within the PSV model is designed to be attached to a single unit of content. But sometimes a unit of content, such as an article, is made up of components that may have their own metadata. The Components block allows publishers to attach metadata to structures within a unit of stored content. This field will likely be used when designing highly complex publishing systems where content management systems and digital asset systems are federated using PSV. See Figure 2.10.

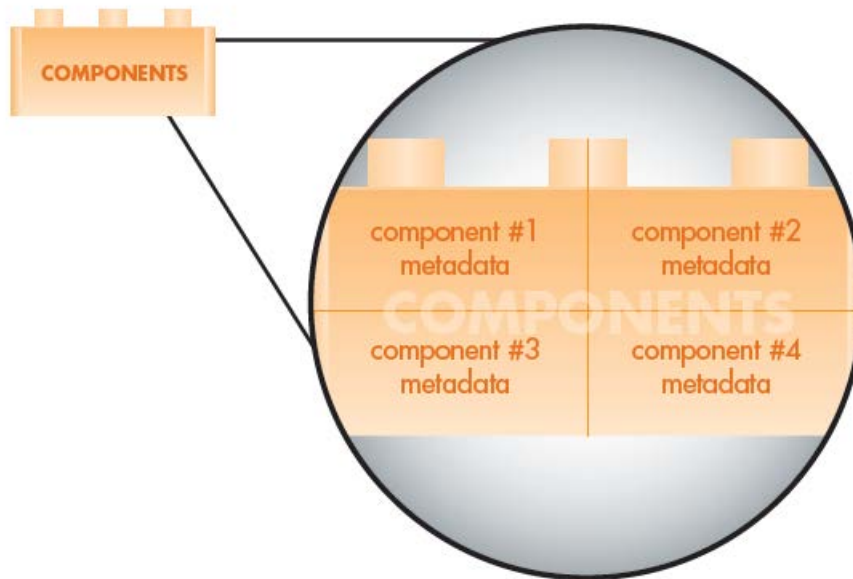


Figure 2.10 Example Components Metadata Fields

3.5 PSV Content Management Schema

In order to assist implementers develop a PSV-based federated content management solution, the nextPub Working Group is providing an XML Schema (XSD) that can serve as the basis for the design of a PSV content repository.

Note: The PSV CM schema is not designed for tagging content. It is provided simply to serve as a basis for the design of a content repository. Metadata building blocks from this schema can be combined with HTML5 by publishers who wish to develop a hybrid PSV metadata and content tagging schema.

3.6 Other PSV Schemas

Because PSV is a flexible framework, it supports many different use case scenarios. A different schema, using the PSV metadata fields and content encoding can be developed for each different use case. In order to assist PSV implementers, the nextPub Working Group is planning to provide a number of XML Schemas (XSDs) to support common use cases including tagging an article and transmitting articles to content aggregators. These PSV sample schemas will be

available from the nextPub website (<http://www.nextpub.org>) and documented in the nextPub PSV Implementation Guide that will be published following the publication of this specification.

3.7 About Advertising

The inclusion of Ad Materials is a design goal of PSV. It is not likely that ads will be stored in the source XML database, ads may be stored in a portal repository. PSV provides us with a starting point for integrating ad materials with source content so it can be delivered to publishing platforms and channels. Note that one of the PSV content types is “advertisement.” The nextPub Working Group is closely collaborating with other groups that are defining advertising metadata so that ad materials can be integrated into the PRISM Source Vocabulary.

3.8 PSV Content Building Blocks

When work began on the nextPub XML Source Specification, the majority of participants favored retaining PRISM/PAM XML as the source format and extending that source to take into account functionality required by the new use cases for content delivered to tablets, eReaders, smart phones and beyond. However, as work progressed and transforms into the two key target delivery formats were considered, it became apparent that we needed to think forward.

With HTML5 gaining favor as an alternative to the OS-based publication apps that we see today on tablets and smart phones, and with the further incentive that EPUB 3 is using HTML5 as the base content format, it began to make sense to bring PSV into as close compliance with the XML Serialization of HTML5 as is possible. After some study, both of the practical issues with using HTML5 as a delivery/display mechanism today and the promise of HTML5 for the future, the decision has been made to move away from PRISM/PAM XML/XHTML as the source and instead move toward using HTML5 to capture content digitally.

Note: nextPub has defined a valid subset of HTML5 for coding PSV content units. While we are recommending some restrictions to HTML5, no extensions have been defined. The goal is to enable any HTML5 browser-based display mechanism to process raw PSV content without plug-ins or customization. This is a major difference between the nextPub strategy and that of EPUB3 that extends HTML5 and requires a conformant EPUB3 reader technology to render EPUB3 content.

3.8.1 Benefits of Employing HTML5

Several benefits can be realized by specifying HTML5 for encoding PSV content:

- HTML5 is a highly flexible, semantic markup that can easily be used to encode a wide variety of content structures using off-the-shelf tools.
- HTML5 can be directly displayed by off-the-shelf HTML5 browsers and processed by HTML5 processor technologies.

- Transforms from PSV into a wide variety of digital delivery/display formats will be much more straightforward because many of these target formats are HTML5 based.
- By using HTML5 to handle rich media as intended, we eliminate the requirement to develop our own rich media models for the XML Source.

3.8.2 Optional HTML5 Head Element

Publishers may use the HTML5 `<head>` structure when capturing digital content. . The HTML5 `<head>` structure includes the title, base, link, script and style elements that add functionality required for proper rendering and display of the elements in the body. PSV recommends against using the HTML5 `<meta>` tag in the head so that all metadata is forced into PSV metadata fields.

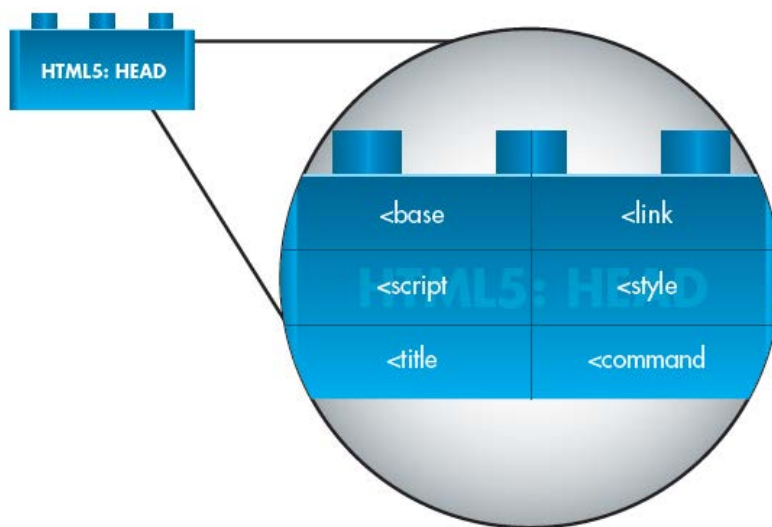


Figure 2.8 Example HTML5 Head Tags

3.8.3 Required HTML5 Body Element

The HTML5 Body is one of the required PSV building blocks. In moving away from the PRISM/PAM XML model to an HTML5-compliant model for coding the body of an article, PSV urges publishers take advantage of the new HTML5 elements that provide semantics that are useful in coding published content. See Figure 2.9.

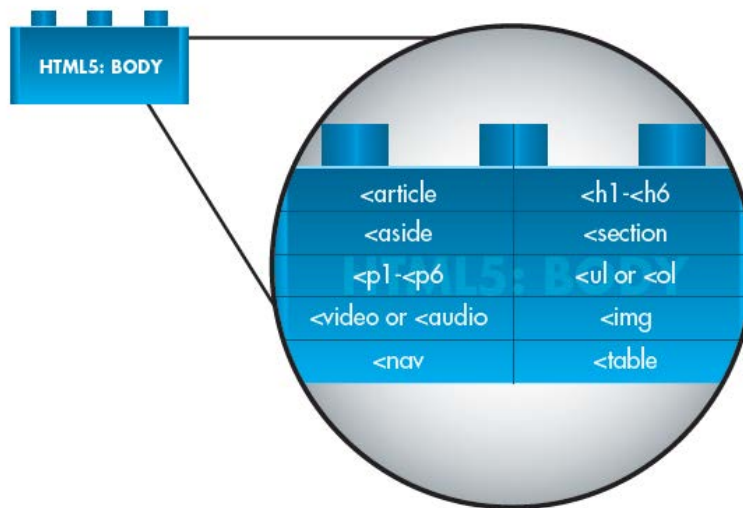


Figure 2.9 Example HTML5 Body Tags

3.8.3.1 Article as the Root Element

PSV specifies that the new HTML5 `<article` tag be used as the root element for any content unit. According to [HTML5], “The `<article` element represents a section of a page that consists of a composition that forms an independent part of a document, page, or site. This could be a forum post, a magazine or newspaper article, a Web log entry, a user-submitted comment, or any other independent item of content.”

3.8.3.2 Additional Semantic Tags

HTML5 has added a number of other semantic tags for publication content. Of particular use for PSV tagging will be the new `<aside`, and `<section` tags. PSV recommends using these tags where appropriate.

Tag	Use
<code><aside</code>	The aside element represents a section of a page that consists of content that is tangentially related to the content around the aside element, and which could be considered separate from that content. Such sections are often represented as sidebars in printed typography.
<code><section</code>	The section element defines sections in a document. Such as chapters, headers, footers, or any other sections of the document.
<code><nav</code>	The nav element tag defines a section or block of navigation links.

3.8.3.3 Figure and Rich Media Tags

HTML5 adds the new figure element as well as tags to embed images, audio and video. A figure is a media block typically presenting *one or more media objects along with ancillary text content*.

This means that the figure is what HTML5 calls a “grouping structure.”

A `<figure>` is not a required to embed rich media objects in an article. If you wish to link a rich media object right into the content stream you may use an `<a>` tag to link to the object, just as you do on a Web page. You can also use ``, `<audio>`, or `<video>` directly in the content stream. Use the figure if you want to include other content associated with the media, such as a caption. Use the figure if you want to group more than a single media object.

3.8.3.4 Adding Semantics Using the HTML5 Class Attribute

To enable semantic content encoding, HTML5 has added a “class” attribute on every element. PSV recommends a number of PRISM semantic classes that you can use to qualify any HTML5 element. Examples include box, caption, dateline, credit, and pull quote.

See the example below:

```
<aside class="prism:pullQuote">
  <p>WE TALK ALL THE TIME ABOUT HOW TO SET THE RIGHT TONE FOR OUR PLAYERS</p>
</aside>
```

4 RELATIONSHIP TO OTHER SPECIFICATIONS

Because there are already so many specifications and standards, IDEAlliance recommends compliance with existing standards and specifications. This Specification recommends the use of certain existing standards, including PRISM 3.0, XML and HTML5.

4.1 Relationship to PRISM

PRISM Source Vocabulary builds upon the foundation of PRISM Specifications and relies on the metadata fields and controlled vocabularies defined by PRISM. PSV defines XML structures for designing a content repository or for tagging source content, but it *does not* define its own metadata fields or controlled vocabularies.

Note: The development of PSV requires the addition of new metadata fields and controlled vocabularies to support PSV functionality. Therefore PRISM 3.0 will be published simultaneously with the publication of PSV 1.0 and will be highly referenced by the PSV Specification. See Figure 3.1.



Figure 3.1 Relationship of PRISM, PAM and nextPub

4.2 Relationship to PAM

PAM is the PRISM Aggregator Message. PAM is an XML tag set built on the foundation of PRISM metadata and controlled vocabularies. The use case for PAM was originally to encode

magazine articles in XML to deliver content to aggregators. While some publishers currently use PAM XML as a content source, that was not the original intent. PAM is an application of PRISM, but PAM and PRISM are not synonymous. PAM is an XML tag set that uses PRISM metadata for a very specific purpose while PRISM remains the core specification for metadata and controlled vocabularies. See Figure 3.1.

Note: Because PSV is not built directly on PAM, The PAM to PSV Guide [PAMPSVGUIDE] has been developed to document the differences between the two XML tagging schemes and how a user might move from PAM encoded content to PSV. This Guide will enable those publishers currently using PAM XML to encode their source content to move to PSV as the XML content source in the future.

4.3 Relationship to HTML5

HTML5 is gaining credibility as a delivery platform for publications, not only on the Web but as an alternative to the publication apps that we see today on tablets and smart phones. With the further incentive that EPUB3 is using HTML5 as the base content format for EPUB3, it makes sense to bring PSV into as close compliance with the XML Serialization of HTML5 as is possible, while still retaining sufficiently rich source semantics. Hence the decision has been made that nextPub XML Source will move from today's PRISM/PAM XML that is based on XHTML to a format which allows for expanded PRISM 3.0 based metadata in the nextPub metadata block while moving to an HTML5-compliant head and body for content encoding.

A *nextPub to HTML5 Transformation Guide* to document the transformation of nextPub XML into HTML5 or delivery to browsers will be added to the nextPub Documentation Set in 2012.

Note: nextPub intends to use the full range of HTML5 along with CSS3 and JavaScript. It places no functional restrictions on HTML5 and makes no extensions. In this way, nextPub differs from the profile of HTML5 specified by EPUB3. See EPUB3 [HTML5 Deviations and Constraints](#) for more information.

4.4 Relationship to EPUB 3.0

Since the beginning, the IDEAlliance nextPub Working Group has worked in collaboration with the IDPF (International Digital Publishing Forum) EPUB3™ Working Group, to design a set of specifications that reference and can be referenced by EPUB to deliver and display magazine, journal, serial, book, newspaper and other rich content for the next generation of eReaders. PSV was designed to complement EPUB3 by providing a compatible content source. PSV defines rich semantic encoding for *source* content while EPUB3, defines a packaging, delivery and destination display format. In some cases PSV content will be aggregated, transformed, formatted and packaged as EPUB3 for delivery to the eReader channel. A Guide to packaging PSV as EPUB3 for delivery and display on eReaders will be made available during 2012.

5 PSV NAMESPACE, DOCUMENTATION AND CM SCHEMA

5.1 The PSV Namespace

Dublin Core, PRISM, PRISM Usage Rights, PRISM Recipe Metadata and other relevant PRISM metadata namespaces along with the new PSV: namespace will be utilized as appropriate. XML structures unique to this specification for content encoding will be given the namespace psv:

The recommended namespace for PSV markup is:

`xmlns:psv="http://prismstandard.org/namespaces/psv/1.0/"`

Note: There is no official schema for HTML5. However, nextPub has developed a schema for the PRISM Source Vocabulary Specification for the purpose of illustrating this specification. For content encoding, this schema is specifically designed to be more restrictive than a true HTML5 schema would be in order to make source content encoding and transformations to delivery channel formats much more straightforward. The design goal was to define a valid HTML5 subset for the content encoding portion of PSV.

5.2 The PSV Version 1.0 Documentation Package

nextPub has developed a series of specifications collectively known as the PRISM Source Vocabulary or PSV. The use case for PSV is to encode semantically rich content for transformation and delivery to any platform. This Specification is made up of a modular documentation package that builds on PRISM 3.0 and HTML5. Over time new modules may be added to the documentation package. The documentation package for the nextPub PRISM Source Vocabulary Specification consists of:

Document	Description
PRISM Source Vocabulary Specification Overview [PSVSO]	The Overview to the PRISM Source Vocabulary provides an introduction and a non-technical overview of the PRISM Source Vocabulary.
PRISM Source Vocabulary Specification [PSVS]	The PRISM Source Vocabulary Specification defines semantically rich for source metadata and content markup that can be transformed and served to a wide variety of output devices including eReaders, mobile tablet devices, smart phones and print.
PRISM Source Vocabulary Markup Specification [PSVMS]	The PSV Markup Specification documents the XML tags in the PSV namespace that are used to encode XML Source Content.
PAM to PSV Guide [PAMPSVGUIDE]	This Guide documents mappings from PAM XML to PSV XML. It is normative only.

5.3 The PRISM Documentation Package

Because PSV is built on PRISM 3.0, there is a close relationship between the two specifications. In fact, access to the PRISM 3.0 Documentation Package is critical to the implementation of PSV. The PRISM 3.0 Documentation Package consists of:

5.3.1 General Documents

This is a set of general or overview documents that apply to PRISM.

Document	Description
PRISM Introduction [PRISMINT]	Overview, background, purpose and scope of PRISM; examples; contains no normative material.
PRISM Compliance [PRISMCOMP]	Describes three profiles of PRISM compliance for content and systems; includes normative material.

5.3.2 PRISM Metadata Specifications

This is the set of documents that outline the prism metadata fields and values by PRISM metadata category. PRISM has modularized its metadata specification by namespace so users may pick those modules that meet their unique business requirements without having to implement the entire PRISM specification.

Document	Description
The PRISM Basic Metadata Specification [PRISMBMS]	Describes the basic metadata elements contained in the PRISM namespace to describe article content; includes normative material.
PRISM Advertising Metadata Specification [PRISMADMS]	Describes advertising metadata elements including those drawn from Ads-ML, GWG and Ad-ID; includes normative material.
The PRISM Subset of Dublin Core Metadata Specification [PRISMDCMS]	Describes the metadata elements from the Dublin Core namespace that are included in PRISM; includes normative material.
<u>The PRISM Image Metadata Specification</u> [PRISMIMS]	Describes the metadata elements contained in the PRISM Metadata for Images Namespace and other related image namespaces, includes normative material.
The PRISM Recipe Metadata Specification [PRISMRMS]	Describes the metadata elements contained in the PRISM Recipe Metadata Namespace, includes normative material
The PRISM Usage Rights Metadata Specification [PRISMURMS]	Describes the metadata elements contained in the PRISM Usage Rights Namespace; includes normative material. This namespace will supersede elements in both the prism: and prl: namespaces in version 3.0 of the specification.

5.3.3 PRISM Controlled Vocabulary Specifications

These modules are new with PRISM 3.0. Previously both the Controlled Vocabulary Markup and the documentation of controlled vocabularies were published as a single document. All controlled vocabularies and their terms are documented in this publication set.

Document	Description
The PRISM Controlled Vocabulary Markup Specification [PRISMCVMS]	Describes the metadata fields in the PRISM Controlled Vocabulary Namespace that can be used to describe a controlled vocabulary. Actual PRISM controlled vocabularies are now placed in the PRISM Controlled Vocabularies Specification [PRISMCVS]
The PRISM Controlled Vocabularies Specification [PRISMCVS]	The PRISM Controlled Vocabularies are now documented in this document.

Appendix A NORMATIVE REFERENCES

A.1 NORMATIVE REFERENCES FOR PRISM AND PSV SPECIFICATIONS

[AAT] Getty Art and Architecture Thesaurus.

http://www.getty.edu/research/conducting_research/vocabularies/aat/

[ADSM-AT] AdsML Ad Ticket 1.0 May 2011, <http://adsm.cnet.se/AdsML-Framework-3.5-R3/AdsMLAdTicket/AdsMLAdTicket-1.0-Specification-AS.pdf>

[DCMI] Dublin Core Metadata Element Set, Version 1.1: Reference Description.

<http://purl.org/dc/documents/rec-dces-19990702.htm>

[DCMI-TERMS] Dublin Core Metadata Terms, 2005-01-10; .

<http://dublincore.org/documents/2005/01/10/dcmi-terms/>

[DOI-HB] Digital Object Identifier Handbook, Version 4.4.1, October 2006,

<http://www.doi.org/hb.html>.

[EPUB3] International Digital Publishing Forum (IDPF), EPUB Version 3.0 2011,

<http://idpf.org/epub/30>

[HTML5] HTML5: A Vocabulary and Associated APIs for HTML and XHTML Working Draft March 2012 <http://www.w3.org/TR/html5/>

[IETF-MIMETYPES] Internet Assigned Numbers Authority (IANA); Internet Media Types.

<http://www.iana.org/assignments/media-types/>

[[IETF RFC 3066](#)], Tags for the Identification of Languages ed. H. Alvestrand. 2001.

<http://www.ietf.org/rfc/rfc3066.txt>

[IMAGEGUIDE] PRISM Working Group 2012, Guide to PRISM Metadata for Images

http://www.prismstandard.org/guides/ImageGuide_3.0.pdf

[INTROPSV] nextPub Working Group, Introduction to the PRISM Source Vocabulary V1.0

http://www.prismstandard.org/specifications/psv/1.0/Intro_PSV.pdf

[IPTC-NEWSML] International Press and Telecommunications Council, NewsML Specification & Documents; <http://www.iptc.org/cms/site/index.html;jsessionid=afhGDa4XwZih?channel=CH0111>

[IPTC-NITF] International Press and Telecommunications Council, News Industry Text Format.

<http://www.iptc.org/cms/site/index.html;jsessionid=afhGDa4XwZih?channel=CH0107>

[ISO-639] ISO 639 - Codes for the representation of names of languages.

<http://www.oasis-open.org/cover/iso639a.html>

[ISO-3166] ISO 3166 - Codes for the representation of names of countries and their subdivisions.

http://www.iso.org/iso/country_codes/iso_3166_code_lists/english_country_names_and_code_elements.htm

[NAICS] North American Industry Classification System; 1997.
<http://www.census.gov/epcd/www/naics.html>

[RFC-3066] H. Alvestrand; Tags for the Identification of Languages; January 2001.
<http://www.ietf.org/rfc/rfc3066.txt>

[PAMGUIDE] PRISM Working Group, 2012, Guide to the PRISM Aggregator Message V. 2.2,
http://www.prismstandard.org/guides/PAMGuide_2.2.pdf

[PAMPVGUIDE] nextPub Working Group, 2012, PAM to PSV Guide V 1.0,
http://www.prismstandard.org/specifications/psv/1.0/PAM_PSV.pdf

[PRISMADMS] PRISM Working Group, 2012, PRISM Advertising Metadata Specification v3.0
http://www.prismstandard.org/specifications/3.0/PRISM_ad_materials_Metadata_3.0.pdf

[PRISMINT] PRISM Working Group, 2012, PRISM Introduction, V 3.0,
http://www.prismstandard.org/specifications/2.1/PRISM_introduction_3.0.pdf

[PRISMBMS] PRISM Working Group, 2012, The PRISM Basic Metadata Specification V3.0,
http://www.prismstandard.org/specifications/3.0/PRISM_Basic_metadata_3.0.pdf

[PRISMCB] PRISM Working Group 2012, PRISM Cookbook, <http://www.prismstandard.org/resources/>

[PRISMCOMP] PRISM Working Group, 2012, PRISM Compliance, V 3.0,
http://www.prismstandard.org/specifications/2.1/PRISM_compliance_3.0.pdf

[PRISMIMS] PRISM Working Group 2012, The PRISM Image Metadata Specification
http://www.prismstandard.org/specifications/3.0/PRISM_image_metadata_3.0.pdf

[PRISMRRMS] PRISM Working Group, 2012, The PRISM Recipe Metadata Specification
http://www.prismstandard.org/specifications/3.0/PRISM_recipe_metadata_3.0.pdf

[PRISMDCMS] PRISM Working Group, 2012, The PRISM Subset of the Dublin Core Namespace V 3.0,
http://www.prismstandard.org/specifications/3.0/PRISM_dublin_core_markup_3.0.pdf

[PRISMURMS] PRISM Working Group, 2012, The PRISM Usage Rights Metadata Specification V3.0,
http://www.prismstandard.org/specifications/3.0/PRISM_usage_rights_metadata_3.0.pdf

[PRISMCVMS] PRISM Working Group, 2012, The PRISM Controlled Vocabulary Markup Specification V 3.0,
http://www.prismstandard.org/specifications/3.0/PRISM_controlled_vocabulary_markup_3.0.pdf

[PRISMCVS] PRISM Working Group, 2012, The PRISM Controlled Vocabularies Specification
http://www.prismstandard.org/specifications/3.0/PRISM_CV_Spec_3.0.pdf

[PSVS] nextPub Working Group, 2012, PRISM Source Vocabulary Specification V1.0
<http://www.prismstandard.org/specifications/psv/1.0/PSV.pdf>

[PSVSO] nextPub Working Group, 2012, PRISM Source Vocabulary Specification Overview V1.0
http://www.prismstandard.org/specifications/psv/1.0/PSV_overview.pdf

[PSVMS] nextPub Working Group, 2012, PRISM Source Vocabulary Markup Specification V 1.0
http://www.prismstandard.org/specifications/psv/1.0/PSV_Markup.htm

[RECIPEGUIDE] PRISM Working Group, 2012, **Guide to PRISM Recipe Metadata and XML Encoding**
http://www.prismstandard.org/guides/RecipeGuide_3.0.pdf

[RIGHTSGUIDE] PRISM Working Group, 2012, **Guide to PRISM Usage Rights**,
http://www.prismstandard.org/guides/RightsGuide_3.0.pdf

[RFC-2119] S. Bradner, **Key words for use in RFCs to Indicate Requirement Level**
<http://www.ietf.org/rfc/rfc2119.txt>

[RFC-2396] IETF, **Uniform Resource Identifiers (URI): Generic Syntax**, Internet RFC 2396.
<http://www.ietf.org/rfc/rfc2396.txt>

[TGN] Getty Thesaurus of Geographic Names.
http://www.getty.edu/research/conducting_research/vocabularies/tgn/

[W3C-DATETIME] Misha Wolf, Charles Wicksteed; **Date and Time Formats**; W3C Note;
<http://www.w3.org/TR/NOTE-datetime.html>

[W3C-RDF] Ora Lassila, Ralph R Swick, **Resource Definition Framework (RDF) Model and Syntax Specification**. <http://www.w3.org/TR/rdf-primer/>

[W3C-RDFS] Dan Brickley, R. V. Guha (eds.), **Resource Description Framework (RDF) Schema Specification 1.0**, 10 February 2004, <http://www.w3.org/TR/rdf-schema/>

[W3C-XML] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen (eds.), **Extensible Markup Language (XML)** <http://www.w3.org/TR/xml/>

[W3C-XML-NS] Tim Bray, Dave Hollander, Andrew Layman (eds.); **Namespaces in XML**.
<http://www.w3.org/TR/xml-names/>

[XMP] Adobe Systems, **XMP Specification**.
http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf

A.2 NON-NORMATIVE REFERENCES FOR PRISM AND PSV SPECIFICATIONS

[ICE] **The Information and Content Exchange (ICE) Protocol**. <http://www.w3.org/TR/NOTE-ice.html> ,
<http://www.icestandard.org>

[ISO-8601] ISO (International Organization for Standardization), **ISO 8601:1988 (E) Data elements and interchange formats - Information interchange - Representation of dates and times**, 1998.
<http://www.iso.ch/cate/d15903.html>

[PRISMIMS] PRISM Working Group, 2012, **The PRISM Inline Markup Specification V3.0**,
http://www.prismstandard.org/specifications/3.0/PRISM_inline_Markup_3.0.pdf

[PRISMPAMMS] PRISM Working Group, 2010, **The PRISM PAM Markup Specification**,
http://www.prismstandard.org/specifications/2.1/PRISM_PAM_Markup_2.1.pdf

[RDF-FRIENDLY] **Making Your XML RDF-Friendly** by John Cowan and Bob DuCharme:
<http://www.xml.com/pub/a/2002/10/30/rdf-friendly.html>

[RDF-SEMANTICS] **RDF Semantics**, Patrick Hayes, Editor, W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> . [Latest version](#) available at <http://www.w3.org/TR/rdf-mt/> .

[XML-SCHEMA1] **XML Schema Part 1: Structures** W3C Recommendation, World Wide Web Consortium, 2 May 2001. This version is <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/> . The [latest version](#) is available at <http://www.w3.org/TR/xmlschema-1/> .

[XML-SCHEMA2] **XML Schema Part 2: Datatypes**, W3C Recommendation, World Wide Web Consortium, 2 May 2001. This version is <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/> . The [latest version](#) is available at <http://www.w3.org/TR/xmlschema-2/> .