Building a Semantic Web of Comics: Publishing Linked Data in HTML/RDFa Using a Comic Book Ontology and Metadata Application Profiles

A thesis submitted to the College of Communication and Information of Kent State University in partial fulfillment of the requirements for the Master of Library and Information Science and Master of Science dual degree program

by

Sean Petiya

December, 2014

Thesis written by

Sean Petiya

B.F.A., Youngstown State University, 2008

M.L.I.S., Kent State University, 2014

M.S., Kent State University, 2014

Approved by

____

Marcia Lei Zeng, Advisor

____

Jeffrey W. Fruit, Director, School of Library and Information Science

____

AnnMarie LeBlanc, Dean, College of Communication and Information

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Definition
AAA	Anyone can say Anything about Anything
AP	Application Profile
ASP	Active Server Pages
CBO	Comic Book Ontology
CBV	Comic Book Vocabulary
CGC	Certified Guaranty Company
CSV	Comma-Separated Values (or Character-Separated Values)
CWA	Closed World Assumption
DCAP	Dublin Core Application Profile
DCMES	Dublin Core Metadata Element Set
DTD	Document Type Definition
FOAF	Friend of a Friend
FRBR	Functional Requirements for Bibliographic Records
GCD	Grand Comics Database
HTML	Hypertext Markup Language
IDE	Integrated Development Environment
IIS	Internet Information Services
JSON	JavaScript Object Notation
KOS	Knowledge Organization Structure
LOD	Linked Open Data
LOV	Linked Open Vocabulary
N3	Notation3
NUNA	Non-Unique Name Assumption
OWA	Open World Assumption
OWL	Web Ontology Language
PHP	Hypertext Preprocessor
RDA	Resource Description Access
RDF	Resource Description Framework
RDFa	Resource Description Framework in Annotations
RDFS	Resource Description Framework Schema
SKOS	Simple Knowledge Organization System
SQL	Structured Query Language
Turtle	Terse RDF Triple Language
VIAF	Virtual International Authority File
XSD	XML Schema Definition
XSL	Extensible Stylesheet Language
XSLT	Extensible Stylesheet Language Transformations

List of Namespaces

Prefix	Namespace
BF	http://bibframe.org/
CBO	http://comicmeta.org/cbo/
CBV	http://comicmeta.org/vocab/
DC	http://purl.org/dc/elements/1.1/
DCTTERMS	http://purl.org/dc/terms/
DCTYPE	http://purl.org/dc/dcmitype/
FRBR	http://purl.org/vocab/frbr/core#
LC	http://id.loc.gov/
OWL	http://www.w3.org/2002/07/owl#
RDF	http://www.w3.org/1999/02/22-rdf-syntax-ns#
RDFA	http://www.w3.org/ns/rdfa#
SCHEMA	http://schema.org/
SKOS	http://www.w3.org/2004/02/skos/core#
TGN	http://vocab.getty.edu/tgn/
VIAF	http://viaf.org/viaf/
VS	http://www.w3.org/2003/06/sw-vocab-status/ns#
XSD	http://www.w3.org/2001/XMLSchema#

Acknowledgements

I would first like to thank my thesis committee, Dr. Karen Gracy and Dr. David Robins, for their feedback and input, and my advisor Dr. Marcia Zeng for her commitment, advice, and direction during the course of this project and in my studies. I am also grateful to all of the wonderful people that I have had the opportunity to meet during my time at Kent State University, and for their inspiration, encouragement, and valuable insights, which have shaped this project and influenced its outcome. And to my friends and family, thank you so much for your patience and understanding, and especially for all of your support.

Chapter I

Introduction

Comics are visual narratives that contain artwork and stories produced by many talented creators, and are enjoyed by readers from a variety of backgrounds and disciplines. Information about the various resources and entities in the comic book universe can be found in a wide range of formats and systems, including those of libraries, archives, and museums, as well as the records of independent research projects and collections. Semantic Web technologies and standards present an opportunity to connect these resources using Linked Data, a set of principles for publishing structured data that enables information to be interlinked between multiple systems. Missing, however, is an open, shared domain ontology or metadata vocabulary, used within the community to identify a comic book at all levels of description and perspective, or describe the specific concepts, entities, and relationships unique to this world. This project presents a case study for the development of a domain ontology for comic books and comic book collections. In the initial phase of the study, a metadata vocabulary and domain model were built from a review of information resources and example materials. A workflow was then designed and used to test the conversion of data from CSV to XML, then to RDF/XML, using the schemata developed. In the second phase, the usability of the vocabulary with regard to publishing Linked Data on the Web is addressed by modularizing the ontology into a set of metadata application profiles (AP) designed to meet the anticipated goals and functional requirements of various user groups, which were identified using preliminary user research techniques that include a review of existing information systems and an analysis of content. Solutions are then discussed for implementing the vocabulary in existing Web content, and a series of examples tested for their ability to generate valid, meaningful RDF data from HTML that is also connected to a variety of Linked Open Data (LOD) resources containing data about comic books and graphic novels.

1.1 Background

1.1.1 History, Collections, and Archives

Once seemingly outcast from the bibliographic universe, the comic book has received a large amount of due attention in recent years as the popularity of the superhero movie genre has generated renewed interest in the medium, and the graphic novel—a very popular format of comic book—continues to make its way to library shelves, becoming embedded within a variety of library collections, services, and programs. Once targeted by authorities, ignored by academia, and while often still the victim of censorship, sentiment towards the comic book as a viable medium for research, scholarship, and pleasure has grown significantly. Research and interest in the comic book is shared across a number of knowledge domains within the humanities, including literary and cultural studies, visual arts, and the field of library and information science. Many libraries include comic books in their special collections, while many museums and archives house collections of comic books and comic art, indicating their status as valid cultural artifacts and historic objects requiring dutiful cataloging, description, and preservation. For example, the Comic Art Collection of the Michigan State University (MSU) Library (comics.lib.msu.edu), with over 200,000 items, including 45,000 foreign comic books, represents one of the original efforts in comics librarianship, a project started in the mid-1970s from a donation of nearly 6,000 comic books (Scott, 1998). The Comic Art Collection at MSU also represents a collaborative indexing effort with the Grand Comics Database (GCD), an international project of volunteers with the ultimate goal of collecting data for "every comic book ever published in every country around the planet" (GCD, 2012b, para. 2). Related efforts are shared with a number of comics research libraries and independent projects maintained by knowledge institutions, scholars, comic artists and writers, and especially fans and collectors of the medium. Many of these research efforts are collected in the MSU Comic Art Collection's "Library Directory" (comics.lib.msu.edu/otherlib.htm) and the *Comics Research Bibliography*, a comprehensive list containing over 29,700 entries (Rhode & Bullough, 2009). Similar in size and scope to the MSU Comic Art Collection, the Library of Congress (LOC) houses the largest comic book collection available to the public containing over 120,000 comic book issues (LOC, 2013). Primarily composed of English-language materials, the collection also includes reprinted Japanese manga, Spanish-language items published in Mexico and the United States, and popular foreign comic book publications (LOC, 2013).

In addition to the bibliographic characteristics and relationships inherent to comic books, and complementing their visual narratives, are the pages of artwork that illustrate their stories. Each page of original art can be considered an individual work, with artwork from popular issues and by leading artists often having a significant monetary value attached to their overall worth (Overstreet & Vaughn, 2013a). These pages may also represent the various stages of production in the creation of comic art, including the sketches, pencils, letters, inks, and colors. An example of an ongoing effort to preserve original comic art is The Jack Kirby Museum and Research Center's Original Art Digital Archive (kirbymuseum.org) of over 3,000 pages of original comic art produced by the museum's namesake, an illustrator whose style is recognized as incredibly influential to the genre of the American superhero comic. Parallel to these collections and research efforts, are the countless personal collections of hobbyists and professional collectors, a community referred to by Serchay as “the serials librarians of the home” (1998, p. 57).

While many hobbyist collections and personal libraries of modern materials often contain similar content, many professional and original owner collections contain items both rare and important to the history of the medium. Significant collections of vintage comic books are considered pedigrees, containing exemplary copies of key issues, or other rare publications. In many cases, these single-owner collections of older materials have been responsible for preserving the work and history of many early creators and publishers from a time when many comic books were often *pulped*, or recycled. One example is the Edgar Church Collection, discovered by comics retailer Chuck Rozanski in 1977, and considered to contain many of the best known copies of Golden Age (1938-1950) comic books (comicbookpedigrees.com). Other collections of Golden Age material represent the provenance for projects digitizing early comic books that have entered the public domain due to copyright expiration.

The comic book has been a recognizable piece of Americana since this Golden Age, and in recent years, like many other forms of literature, has made its way to a digital format which can now be read on smartphones, tablets, and other devices. This improved accessibility has helped renew interest in the medium (O’Leary, 2013), but perhaps renewing interest in the comic book more than anything else has been the influx of superhero movies appearing at the box office, with films starring characters like Batman, Spider-Man, and Captain America grossing record worldwide earnings (Pyles, 2013; boxofficemojo.com). However, before superhero films were able to grace the screens of movie theaters, or the original tales of these heroes’ deeds could be read on a smartphone, their stories were first printed in four colors across the pages of comic books.

1.1.2 Visual Form and Format

The comic book represents a visual narrative, or a work of sequential art, a concept defined by legendary illustrator Will Eisner as “an art and literary form that deals with the arrangement of pictures or images and words to narrate a story or dramatize an idea” (2008, p. xi). In the seminal work *Understanding Comics*, author Scott McCloud builds upon Eisner’s definition and offers a refined description that reads: “juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or produce an aesthetic response in the viewer” (1994, p. 9). McCloud also stresses the importance of *deliberate sequence* in this passage and draws a distinction between comics and animated cartoons, noting the importance of visual composition to comics, stating “space does for comics what time does for film” (1994, p. 7). This narrative unfolds across a sequence of juxtaposed panels that illustrate the story, word balloons that hold the speech of the characters, and captions that narrate the tale. Each page containing this story within its borders represents the combined creative effort of one or more writers, illustrators, inkers, colorists, letterers, and often many other contributors, involved in the production of a single comic book issue.

In its original magazine format, the published comic book appears as a serialized periodical on newsstands, in bookstores, and in specialty shops, often called *comic book stores*. Each serialized comic is published as part of a series, or title, many of which are divided into multiple volumes. Many comic book titles are published monthly, while others may be published bi-weekly, yearly (an annual), or even just once (a one-shot or limited edition). Often the title, volume numbers, and issue sequence of many series will change, combine, or restart, a reality of comics material which William Fee describes as a “serials cataloger’s nightmare” (2008, p. 176). In addition to its magazine (or “floppy”) format, a run of issues in a series may also be published as a graphic novel, a book-length form made popular in the 1980s, often containing painted panels and mature stories, but now often used as a term to simply describe "a collection of comic book material in book form" (Fagan & Fagan, 2011, p. 5). This collection of comics material can also take the form of a trade paperback or a hardcover, both of which commonly collect multiple issues of a single volume, or an entire storyline. This storyline, or story arc, is a single narrative told across multiple issues within a title's ongoing continuity (Griffin, 1998), and may appear in parts across multiple titles. Larger collected editions may be referred to as *omnibuses* and can collect entire portions of a single volume or series, sometimes containing over 1,000 pages of reprinted comic book material. Many of these editions also contain accompanying materials (Serchay, 1998), such as creator commentary, editorial content, or historic information about related works, such as toys, movies, or videogames, and may even make modifications to the pages of the reprinted work occasionally adding whole panels, or adding and deleting text. Comic books, in any format, are published as a variety of manifestations, including: *variants*, which can be noted by any change to the original production, including the cover, price, or printing; *reprints*, an additional printing or re-release of an issue; and *editions*, which may reference a special or limited edition, but can also refer to a direct market or newsstand edition, the latter of which for a period of time referred to the fact that the comic book was delivered to a newsstand rather than a direct market retailer (Serchay, 1998; Duncan & Smith, 2013).

Many comics are translated and redistributed in different countries, where they are often given a new title, as well as different volume and issue number assignments. While European and Spanish comic books have their own unique histories, they share many close similarities in regard to content, form, and publication format with the American comic book, making the translation and distribution of these comics from one continent to the other very common

(Lefèvre, 2010). In contrast, Japanese comics or *manga*, have a distinctive creative style, publication format, and nature of content that distinguishes it from the American and European comic book (O'Nale, 2010). McCloud identifies distinctions in the visual vocabulary and storytelling techniques of manga, compared to that of American comic books, especially with regard to reader participation (McCloud, 2006). These distinctions regarding format and approach are important, as noted by Duncan and Smith in a discussion on differentiating between similar comic formats, such as the comic *strip* and the comic *book* (2013). The authors suggest four perspectives through which to distinguish variations of the medium, including: (a) production, (b) distribution, (c) art form, and (d) perspective (2013, p. 5). However, despite the formal characteristics or cultural perspectives that may appear to separate these publication formats from one another, they are all *comics* and often find their way from the comic book shop, book store, or library, to the hands and homes of eager readers, where they may even become part of a comic book collection or personal library.

1.2.3 Hobbyists and Researchers

Hobbyist comic book collectors may collect for fun, profit, or even competition. Some collect appearances and titles featuring their favorite character, while others may focus on acquiring complete runs of a series, or on acquiring all of the comics published in a particular year or even by a specific publisher. Collectors typically store comic books in archival bags and keep each item in a box designed specifically for storing comics. Bags are available in varying degrees of archival quality material, including mylar, polyethylene, and polypropylene, with the latter being the least expensive and most common (Cole, 2013). Collectors concerned with the long-term preservation of their collection may choose to store their comics in an acid-free box specifically designed to hold comic books. These boxes are commonly referred to as a *short box*, holding about 150 items, or a *long box*, which holds about 300 items “bagged and boarded” (Cole, 2013). Condition is of significant concern for many collectors and any physical imperfection including rips, tears, marks, stamps, and stains are often noted with the appearance of each one detracting from the visual appeal and value of a comic. Providing reference for determining the physical condition of an item is the *Overstreet Comic Book Price Guide*’s “Grading Definitions”, an illustrated version of which can be found in the companion guide to collecting comics (see Overstreet, 2013; see also Overstreet & Vaughn, 2013b). Collectors owning high-grade, valuable, or rare materials may choose to have those items professionally graded by a third-party service, such as the Certified Guarantee Company (CGC). After a thorough inspection and professional assessment of condition, the guaranty service encapsulates the item in a plastic container and assigns a numeric grade. This grade corresponds with standard grading definitions and represents a certified summary of the item’s physical condition. This process often increases the market value of an item significantly, but affords the buyer in an online marketplace or auction confidence in the condition of their purchase. The population, or total number, of graded items combined with known circulation data of an issue, may aid in determining scarcity, a factor affecting the price of comic book back issues (Duncan & Smith, 2013). Collectors catalog their collections using spreadsheets, custom databases, collection management websites, mobile apps, desktop software applications, or even simply by using pen and paper.

The dedication of comic book collectors to the medium and hobby has not only generated significant contributions to the field of comics research with regard to librarianship and scholarship, it has also generated a large number of personal and independent research projects that each represent unique and significant contributions to the body of available information about comics. Much of this knowledge would have been lost resulting from the deaths of early creators, artists, and other figures within the comic book industry, and the dissolution of many original publishing companies. Projects like the Grand Comics Database (GCD) (comics.org), an international effort to build a detailed index of the world's comic books; Chomichron: The Comics Chronicles (comichron.com), a research project collecting comic book sales figures and circulation data; and both The Digital Comic Museum (digitalcomicmuseum.com) and Comic Book Plus (comicbookplus.com), providers of digitized, public domain comic books, represent just some of the many ongoing comics research efforts.

While the background of the comic book presented certainly describes what can be considered a cataloging "nightmare" what it also represents is a unique opportunity to explore the complex relationships that exist between the various dimensions of the comic book as a bibliographic publication, document, art object, visual sequence, story, and collection item in a variety of systems, catalogs, databases, and indexes. Making the effort particularly rewarding is the expansive, diverse, and collaborative user community, which extends to numerous organizations, fields of study, and into the homes and businesses of many collectors and professionals, with an innate ability, need, and desire to catalog and organize their collections.

1.2 Problem Statement

Research and data about comic books can be found in a variety of systems and formats, including the records of library catalogs and finding aids, personal collection records and management systems, in various independent research projects and databases, and in a large number of print resources, handbooks, and indexes. While the communities responsible for these systems and resources have developed their own systematic routines, standards, and best practices in collecting, cataloging, and describing comic books within their collections—there are limited options available for librarians, collectors, researchers, and historians to share their information and knowledge, or implement bibliographic control over their data without compromise. Tarulli describes the difficulty of representing graphic novels using existing library cataloging practices noting that "[…] the classification of graphic novels, and their subject content [does] not pull together relationships and similarities" (2010, p. 219). Some of these relationships include those that exist between storylines, characters, and other similar or related content from the same publisher. The author, pointing out that fans of the genre are often dismayed by this shortcoming of library catalogs, suggests that the descriptive relationships provided by the Resource Description and Access (RDA) cataloging standard, and implementation of social cataloging or next-generation catalogs will improve the ability of patrons to discover graphic novels in the library (Tarulli, 2010). The collaborative cataloging relationship between the Michigan State University (MSU) Library and the Grand Comics Database (GCD), described by Scott (1998), demonstrates how libraries and other knowledge institutions can both share and enrich their data through cooperation with related research projects and user communities. Furthermore, participation in this collaborative environment and exchange of data can be extended to publishers, creators, and retailers—and as interest in original comic art and its history continues to grow—to museums, archives, and galleries as well. However, despite the common elements of data and interest in different aspects and characteristics of the comic book shared by these knowledge institutions, communities, and users, no specific metadata standard is available that allows these identified participants within the comic book domain to share, combine, link, or describe their data at all necessary levels of description, or with the specificity the domain requires. Despite this lack of standardization, the growing "Web of Data" represents an opportunity for this large group of participants to share and exchange data about comic book resources, enhancing and preserving the efforts of librarians, researchers, scholars, historians, creators, and collectors alike who have contributed to the history, acceptance, and proliferation of the medium.

The problem addressed by this study is the development of an ontology for the description of comic books and comic book collections in the Semantic Web, which can also function as a metadata vocabulary for sharing, exchanging, and enhancing records between distinguished participants and communities within the comic book domain, including publishers, retailers, collectors, libraries, and researchers. This study takes a user-centric approach to the development of the ontology and supporting schemata, by focusing on the needs and requirements of individual user groups identified through an analysis of existing information systems and content. It seeks to extend the usability of the vocabulary through the development of a series of metadata application profiles (AP) guided by data-driven user personas representing each participant group, and constructed from a review of existing systems and an analysis of content. The hope is that the effort may also serve as a case study for better enabling community and user participation in the open exchange of data on the Web by presenting a framework for enhancing the usability of the tools, structures, workflows, and methods used to participate in that exchange. The nature of the comic book itself, as a bibliographic resource, document, art object, and collection item—and commitment of the wider community to its preservation and detailed description—makes the domain a unique candidate for studying metadata description, quality, and interoperability, as well as for identifying methods to better enable and enhance user participation in the creation and publication of that data, especially using Web vocabularies in a Linked Data environment.

The process of publishing Linked Data makes information that would otherwise be restricted by limited access available for sharing, reuse, and repurposing to a variety of user communities for research and application development (see linkeddata.org). This process of information sharing should be made as easy and as simple as possible in order to best achieve the vision of a Semantic Web, a "Web of Data" in which both humans and machines can better understand, connect, and relate information (Berners-Lee et al., 2001). Structured data, produced through the implementation of metadata vocabularies and standards, lies at the heart of the Linked Data publication process (W3C, 2013a). While knowledge institutions, such as libraries, archives, and museums, often publish information of the highest quality and caliber, this data may still be incomplete. Extending participation in this description and exchange of information to specialist user communities, independent researchers, and other knowledge domains expands the "collective knowledge" available about a resource (Gruber, 2008). Despite user data often existing in a variety of formats, encodings, and at various levels of quality and consistency, user participation in the cataloging and indexing of data is crucial for the continued growth, maintenance, and overall relevancy of a dataset, as well as to the transfer and expansion of knowledge within a domain. Encouraging user participation in a Linked Data environment requires tools that enhance the usability of implementing a vocabulary, while abstracting the concepts of the Semantic Web and the complexities of underlying knowledge organization structures (KOS), better enabling developers and content authors to focus on preparing and publishing their data, and building great experiences.

In summary, this study addresses the following problems: there is currently (a) no open, shared metadata standard for comics; (b) no domain ontology or comics-specific Web vocabulary; (c) no existing model that comprehensively describes all dimensions of a comic (i.e., as a publication, document, visual sequence, story, product, and collection item); and finally, that (d) most community created data, research, and content is contained in existing hypertext (HTML) information systems, websites, databases, catalogs, etc.

1.3 Objectives

The primary objective of this study is to produce a domain ontology for comic books and comic book collections, and present options for utilizing knowledge organization structures (KOS), specifically metadata vocabularies and related application profiles (AP), to better enable the publication of structured data by groups of users within a community. The study seeks to meet this objective through a case study for comic books, a material rich with multiple levels of bibliographic description, complex relationships and responsibilities, and a continually growing presence in libraries, archives, museums, and academics—in addition to being both a hobby and profession with its own inherent cataloging practices, standards, and requirements shared by collectors, retailers, and publishers, among a variety of other participants. The study also aims to extend the usability and usage of the ontology by designing modular APs for selected user groups informed by a review of existing systems and an analysis of content, which are intended to guide the publication of Linked Data inline using HTML/RDFa. The assumption is that by developing modular APs focused on how specific groups of users within a community publish data, the usability of a detailed, domain ontology can be improved, better enabling users to participate in an open exchange of information. This exchange represents not only the transfer of records between systems, but also the possibility of incorporating detailed, specialist information alongside that of knowledge institutions, increasing the amount and quality of data available to all participants. Within the domain of comic books, a world made increasingly accessible through digital comics, graphic novels, manga, and superhero movies, it may just be historians, researchers, fans, and other champions of the genre that hold the most complete and accurate descriptive data about these popular materials.

Chapter II

Review of Literature and Related Projects

2.1 Modeling Comic Book Data

There has been no shortage of efforts in modeling and representing comic book data in closed software systems, on the Web, or in the classroom. They are fun, popular, and accessible materials enjoyed by many. Similarly, there has been no shortage of efforts to create metadata and markup languages for comics. Much of the effort in developing metadata for comic books has focused on the digital format of the medium, often including markup and semantics that exclusively describes components of web comics or digital comic books. There are currently no open, publically available, deferenceable Web vocabularies (or ontologies) designed specifically for comic books and comic book collections. However, as the digital comic has grown into an ever more popular format, digital comic book XML schemas and markup languages have become necessary, designed to meet the needs and requirements of comic book reading/viewing applications. Notable efforts include: (a) the Advanced Comic Book Format (ACBF), (b) the Comic Metadata (CoMet) Format, (c) ComicsML, and (d) the Comic Book Markup Language (CBML). Each specification describes digital comics to different degrees of granularity, and represents various levels of ongoing development and commitment by the author. However, the researcher was unable to locate any large, publically available datasets that had implemented any of the vocabularies listed above, thus the proceeding discussion utilized each specification's XML Schema Definition (XSD), accompanying documentation, or example records as provided by their respective authors for reference and study.

2.1.1 Advanced Comic Book Format (ACBF)

The Advanced Comic Book Format (ACBF) is described as a "distribution and interchange format for digital comic books" (Pastierovic, n.d., para. 1). It is a detailed schema containing a hierarchy of elements, and through the use of `xsd:enumeration` elements defines a set of controlled term lists within its XSD. This particular specification also addresses the comic book document structure and visual sequence, including individual layers of digital text. ACBF can be used alongside other technical metadata, such as that found in Comic Book Zip (CBZ) files and other archive formats. The author of ACBF maintains a suite of tools for producing ACBF metadata, as well as links to digital comic books that have been converted to the format.

2.1.2 Comic Metadata Format (CoMet)

The Comic Metadata Format (CoMet) is an XML specification for digital comics and an attempt to "standardize the representation of information used to describe comic books" (DenVog, n.d., para. 1). The author of CoMet also publishes a digital comic book reading application titled Comic Viewer, which supports the format. CoMet contains a total of 28 elements, many of which use common element names found in existing specifications like the Dublin Core Metadata Element Set (DCMES). This list of elements in CoMet is non-hierarchical, or flat, with only one required element, and nine that are repeatable. While referenced in the specification, there is no accessible XSD available for validation. However, according to its documentation, both issue and volume are constrained to positive integer values, with most remaining elements being identified as string values. The CoMet specification document also makes no distinction between digital or print comics, and of the formats reviewed appears to have the widest applicability.

2.1.3 ComicsML

ComicsML is identified as an "XML-based markup language for digital comics" (McIntosh, 2011, para. 1). The project represents one of the original efforts to create a markup language for digital comics material, started in 2001 by Jason McIntosh, and believed to have the potential to "help digital comics assert their value as online resources and as art forms" (2001, para. 1). ComicsML is available as a Document Type Definition (DTD), which is a document that declares the valid markup for an XML record implementing the specification. Unfortunately, the project is no longer in development.

2.1.4 Comic Book Markup Language (CBML)

One unique effort that focuses on encoding the documentary aspects, structure, and content of comic books is the Comic Book Markup Language (CBML), an "XML vocabulary for encoding multiform documents that are variously called comics, comic books, and 'graphic novels' […]" (Walsh, 2012a, para. 1). CBML extends the Text Encoding Initiative (TEI) Guidelines for Electronic Text Encoding and Interchange (an XML schema), through the addition of comic book specific elements, including *panel*, *balloon*, and *caption* (Walsh, 2012a). CBML is described as focusing on the "overall structure of the document, textual content, and metadata," noting that "significant work remains to be done giving similar attention to the pictorial dimensions of the comic book" (2012b, para. 62). As an extension of TEI, CBML is a very granular and expressive encoding which allows for detailed descriptions, as indicated by example markup (see Walsh, 2012b), with significant potential for providing accessibility to rare comics material, especially for libraries and archives seeking to digitize and encode their collections.

2.1.5 Summary of Existing Schemas and Data Models

The overview presented above describes the most commonly referenced, independently published XML schemas for digital comic books, each representing various levels of commitment and maintenance by the author, and implementation by the community. Impossible to include in the list are the various local data models and metadata schemas used by comic book viewers, collection management applications, and custom schemas that users have developed to manage their own collections. Because many of these XML schemas describe the presentation or structure of a digital comic they focus heavily on the document and its visual sequence (i.e., the panels, word balloons, and captions), and much less on the container or the object itself, especially with regard to resource discovery and organization. However, the typical data model for a comic book object (print or digital) includes a number of common elements, such as *series title*, *volume number*, and *issue number*. While ACBF, CoMet, ComicsML, and CBML each incorporate many similar entities and attributes common to the modeling and description of a comic book, they focus on the digital representation of the object and its content, and may not be specific or extensible enough to describe many of the nuances and details necessary for accurately representing and cataloging printed comic books. Currently, an open, dereferenceable metadata vocabulary designed for the representation of printed comic books and comic book collections as they exist in stores, libraries, archives, museums, or in the homes of collectors, is not available.

2.2 Semantic Web and Linked Data

The Semantic Web is a "Web of Data" in which meaningful links are made between information resources using a common framework of technologies and standards that allow data to be described, shared, linked, and reused between applications (W3C, 2012a). In contrast, the hypertext Web is a network of linked documents connected through non-descript links containing no information about the nature of the link or the relationships between the documents. The language encoding these documents, Hypertext Markup Language (HTML), primarily defines how a document should be rendered visually to a user when it is accessed in a web browser. A hypertext document contains markup pertaining to both its structure and additional components, such as headings, paragraphs, and lists. However, the ability of machines to automatically interpret this information, or derive meaning and establish automated connections between documents and other resources, is significantly limited (Antoniou, Groth, Harmelen, & Hoekstra, 2012). The notion of a Semantic Web, or "Web of Data", adds meaning to the links between information resources, creating more purposeful connections and better facilitating resource discovery in a networked environment (Berners-Lee, Hendler, & Lassila, 2001). In the W3C "Semantic Web Frequently Asked Questions" document, goals described for the Semantic Web include decoupling "siloed" data from applications, creating a framework for sharing structured data across enterprises and communities, and building tools to automatically reveal new relationships within that data (W3C, 2009). The FAQ goes on to describe additional areas of applicability, which include data integration, resource discovery, classification, and cataloging, among many others (W3C, 2009). In addition to institutions that may be exploring the benefits of publishing structured data, the utility of the Semantic Web described by its goals and applicability also has obvious implications for content authors, publishers, and webmasters looking to enhance the visibility of their data, as well as for users of software applications and services that wish to share, mix, combine, or enhance the data contained in those packages.

2.2.1 Semantic Web Framework

The World Wide Web Consortium (W3C) maintains many of the standards composing the Semantic Web framework, or the Semantic Web Stack, often referred to as a "layer cake". This stack of technologies, as originally envisioned by Tim Berners-Lee, contains a series of layers that compose the architecture of the Semantic Web, including: identifiers, character sets, syntax, data interchange, taxonomies, ontologies, rules, querying, unifying logic, proof, trust, and cryptography, all supporting a top layer that consists of the user interface (Berners-Lee et al., 2001; Wikipedia, 2013). Additionally, there are a number of fundamental concepts important to the functionality of the Semantic Web, and of significance to this study, including: (a) The AAA principle, or "Anyone can say Anything about Any topic"; (b) the Open-World Assumption (OWA) vs. Closed-World Assumption (CWA), implying that data can either be assumed to be incomplete or complete, respectively; and (c) the Non-Unique Name Assumption (NUNA), which states that "the same entity could be known by more than one name" (Allemang & Hendler, 2007, p. 13). To summarize: when viewed together, these concepts imply that (a) multiple systems can provide information about any resource (AAA); (b) the total information available about any given resource is unknown (OWA); and (c) a resource may be referenced using many identifiers from multiple systems (NUNA).

The primary difference between the Semantic Web technology stack and that of the hypertext Web, although they share many of the same components and protocols, is the ability of the Semantic Web to convey meaning about the information it is describing using Web vocabularies, or ontologies (W3C, 2013a). The structured data defined using common vocabularies is machine-processable, and can automate the discovery of links between systems and resources. The principles and practices describing this process are often referred to as Linked Data.

2.2.2 Publishing Linked Data

What composes the Semantic Web are the links between data, or Linked Data (Berners-Lee, 2009). Linked Data refers to the usage of tools and standards from the Semantic Web stack to publish structured data. It also refers to the principles designed to better facilitate the exchange of that data between information systems, and to establish semantic links between data from different sources (Bizer, Heath, & Berners-Lee, 2009; Berners-Lee, 2009). These principles include: (a) using URIs for the names of things; (b) using HTTP URIs so that people can look up those names; (c) providing useful information using standards; and (d) including links to other URIs so that people can discover more things (Berners-Lee, 2009). The process of creating Linked Data involves using common models and standards, like the Resource Description Framework (RDF), to describe the relationships between resources (W3C, 2014a). To build these relationships, the usage of standardized vocabularies, or Linked Open Vocabularies (LOV), is encouraged, which also allows data publishers to participate in the "Linked Data Cloud" by linking to other resources and datasets using shared terms and identifiers (W3C, 2014a). The shape of this cloud—its network of connections—is built by interlinking resources between various published datasets (W3C, 2014a). These links produced by multiple participants add to the total amount of knowledge available about a given resource, effectively expanding the "global graph" describing it, with no contribution of information considered too small (Baker et al., 2011). This process of publishing and linking structured data allows data to be discovered and reused (Bizer et al., 2009), as well as expanded, elaborated upon, and enhanced by other participants in the Linked Data ecosystem. The utilization of semantic descriptions and relationships, and the application of complex knowledge organization structures (KOS) like Web Ontology Language (OWL) ontologies to data in this cloud allows machines to "understand" the meaning of the relationships by making inferences about the content and then interlinking that data to new information automatically, uncovering connections that previously did not exist (Berners-Lee, 2001; W3C, 2013a).

In an analysis of how structured data is published on the Web, Pohorec, Zorman, and Kokol identify two approaches: (a) *inline*, as part of an HTML document, or (b) *parallel* (2013). Structured data published in parallel, as a separate machine-processable collection of RDF data, is a common way of making large datasets available as Linked Data. However, as the authors note in their study, RDF data published using this approach is often not included in traditional search engines, and does not have the same visibility as HTML content that has been crawled and indexed by search services (Pohorec et al., 2013). An alternative approach is to publish structured data by adding semantic annotations to existing HTML content using extensions, like Microdata or Resource Description Framework in Attributes (RDFa). Linked Data principles are syntax independent, which makes the technology compatible with a variety of markup languages and serialization formats.

2.2.3 Markup Languages, Models, and Syntaxes

2.2.3.1 Extensible Markup Language (XML)

XML is a markup language used to describe the structure of a data document and is a popular record format for sharing structured data between information systems. The XML syntax is self-describing, which enhances the readability of the XML record as well as the usability, both in terms of being able to read and understand the contents of an XML record, and with regard to authoring a set of records (W3C, 2010). A typical XML document (Figure 1) contains a root node, and a set of elements. XML elements may include child elements, as well as additional attributes that further describe the data.

```
<Comic>
  <seriesTitle type="alternate">
    The X-Men (1963)
  </seriesTitle>
</Comic>
```

Figure 1. An XML document.

Used in conjunction with XML, XML Schema is a constraint language that provides rules for an XML document regarding what elements can be used in a record, and the type of value those elements can contain. An XML record can be validated using an XML Schema Definition (XSD), making the constraint language an ideal tool for validating the quality of information in a setting where multiple systems are exchanging or consuming data. Many established metadata vocabularies, including Dublin Core (DC) and Visual Resource Association (VRA) Core, provide an XML Schema document that can be used to validate XML records authored using their associated metadata element sets.

XML records declare and define namespaces to avoid ambiguity and provide identification for vocabulary terms. This process enables the terms defined to be reused in other records or schemas. An XML document using a namespace is considered to be *qualified* (Figure 2), and the elements within that document can be identified using the appropriate namespace.

```
<cbo:Comic xmlns:cbo="http://comicmeta.org/cbo/">
  <cbo:seriesTitle type="alternate">
    The X-Men (1963)
  </cbo:seriesTitle>
</cbo:Comic>
```

Figure 2. A qualified XML document.

While XML remains a popular and widely used format for describing and exchanging data, an XML schema embodies a Closed World Assumption (CWA), or the assumption that the data record is complete, meaning it is compliant with the design set forth by the schema. An alternative approach to resource description where not all of the available or possible information may be known, and thus not completely identified by any one schema, is the Resource Description Framework (RDF).

2.2.3.2 Resource Description Framework (RDF)

The Resource Description Framework (RDF) is a framework that can be used to describe a variety of resource types, documents, or concepts (W3C, 2014c). An RDF document is composed of a set of *triples*, which represent a series of statements about the resource being described. The structure of these statements is very similar to that of a sentence. An RDF triple is composed of: (a) a *subject*, (b) a *predicate*, and (c) an *object* (Figure 3).

Comic (subject) → seriesTitle (predicate) → “The X-Men” (object)

Figure 3. An RDF triple.

RDF data can be encoded using a variety of syntax languages. One of the most common is XML, or RDF/XML (Figure 4), which also uses namespaces to identify and disambiguate vocabularies.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:cbo="http://comicmeta.org/cbo/">
  <rdf:Description rdf:type="cbo:Comic" rdf:about="#XMen">
    <cbo:seriesTitle>
       The X-Men
    </cbo:seriesTitle>
  </rdf:Description>
</rdf:RDF>
```

Figure 4. RDF data encoded in RDF/XML.

The encoding of RDF data is not limited to XML and many popular alternatives exist, including the Terse RDF Triple Language (Turtle), which is a significantly more readable format used for representing RDF triples (Figure 5).

```
@prefix cbo: <http://comicmeta.org/cbo/> .
<#XMen>
       a cbo:Comic ;
       cbo:seriesTitle “The X-Men” .
```

Figure 5. RDF data encoded in Turtle.

Finally, the object in an RDF triple is not restricted to literal values, and can also reference or link to another resource using a Uniform Resource Identifier (URI).

A set of linked RDF triples is commonly referred to as a *graph*, which represents the connections between objects or resources, as well as the information that describes them (W3C, 2014c). This model is distributed, and multiple participants can make statements about a resource contributing to a *global graph*, which represents all knowledge and information available about that particular resource (Antoniou et al., 2012). This notion of a graph of information embodies an Open World Assumption (OWA), in which data *cannot* be assumed to be complete, because “there is always more information that could be known” (Allemang &

Hendler, 2007, p.13). Similar to the hypertext Web, in which any participant can create a link between documents, the Semantic Web extends this same sense of universality by expanding the collective knowledge available about a given resource through statements contributed by a variety of participants, both academic and commercial (Berners-Lee et al., 2001). The usage of common metadata vocabularies enables participants in this environment to make a contribution to the graph by publishing data about any resource, and by defining additional relationships between other resources. Web vocabularies defined using RDF and RDF Schema (RDFS), a semantic extension of RDF, define the concepts and relationships that connect these resources in the Semantic Web (W3C, 2013a), and can be used to embed structured data in HTML.

2.2.3.3 Hypertext Markup Language (HTML)

HTML5 is the most recent version of the HTML standard, and includes an additional set of markup elements that provide enhanced semantic details about the structure of a document on the Web. In previous versions, the markup of regions describing the content of an HTML document were not descriptive. Existing elements like `<p>`, for a paragraph of text, or `<div>`, for a division of the document, are unable to convey information about the type of content those regions contain. HTML5 provides enhanced, semantic elements like `<header>`, `<article>`, `<aside>`, `<section>`, and `<footer>` allowing search engines to determine the relevancy of content within a document more accurately, and giving machine readers and browser plugins the ability to better render that same content to the user based on the markup (Ronallo, 2012). In addition to these new components, the HTML5 specification includes the Microdata semantic markup syntax, which gives webmasters the ability to publish structured data by embedding it into existing HTML content (Ronallo, 2012).

2.2.3.4 Microdata

Microdata is a syntax composed of a series of name-value pairs describing groups of content appearing within an HTML document (W3C, 2013b). The groups are referred to as *items* and the name-value pairs as *properties*, the values of which are conveyed through additions to existing elements using `itemscope`, `itemtype`, `itemid`, `itemprop`, and `itemref` attributes (W3C, 2013b) (Figure 6).

```
<section id="ComicBook" itemscope>
  <h1 itemprop="seriesTitle">The X-Men</h1>
  <div>
    Publisher:
    <span itemprop="publisherName">
      Marvel Comics
    </span>
  </div>
</section>
```

Figure 6. Microdata markup added to HTML5 content.

A popular method of publishing structured data using the Microdata syntax is to implement terms from a Web vocabulary, like schema.org, to enhance the semantic description of the content. A combined effort from major search engines Bing, Google, and Yandex, schema.org represents a broad, shared vocabulary that can be used to publish structured data and enhance the visibility of content for search engines and other applications (Ronallo, 2012). In the example below (

Figure 7), an HTML document using Microdata links content to the schema.org vocabulary through the `itemtype` property.

```
<section id="ComicBook" itemscope itemtype="http://schema.org/Periodical">
  <h1 itemprop="seriesTitle">The X-Men</h1>
  <div itemscope itemtype="http://schema.org/Organization">
    Publisher:
    <span itemprop="publisherName">
      Marvel Comics
```

```
    </span>
  </div>
</section>
```

Figure 7. Microdata markup added to HTML5 content using schema.org.

In addition to enhanced search engine visibility and connection to a larger Web of data, the publication of structured data enables the display of *rich snippets*, an improved display of information in a user interface, particularly in a search engine results page (Ronallo, 2012; Google, 2013). While the implementation of Microdata is a popular method for adding semantic annotations to existing HTML content, an alternative and more robust approach is the use of RDFa.

2.2.3.5 Resource Description Framework in Attributes (RDFa)

Resource Description Framework in Attributes (RDFa), is a syntax for embedding structured data in HTML documents (W3C, 2013c). It can be used to implement components from multiple vocabularies, as well as custom metadata elements (Google, 2013). The RDFa syntax is compatible with a variety of markup languages, including HTML5, HTML4, and XHTML, as well as XML and SVG (rdfa.info). Embedding structured data using RDFa improves the visual display of information through machine-readable “hints” (W3C, 2013b), and improves the user experience by providing enhanced functionality, allowing content to be shared, transferred, and more accurately searched (W3C, 2013d). Additionally, RDFa is supported by schema.org, as well as Google’s Rich Snippets and Custom Search technologies (schema.org, n.d.; Google, 2013), which gives webmasters deploying RDFa similar benefits to that of using Microdata.

RDFa is available in two versions, RDFa Core and RDFa Lite, the latter of which represents a simplified, minimal subset of attributes required for adding semantic markup to content (W3C, 2012b). These attributes include `vocab`, `typeof`, `property`, `resource`, and `prefix`, and can be added directly to existing elements in a document (Figure 8).

```
<section id="ComicBook" vocab="http://comicmeta.org/cbo/" typeof="Comic">
  <div property="series" resource="#XMen_1963" typeof="Series">
    <h1 property="seriesTitle">The X-Men</h1>
    <div property="publisher" typeof="Publisher" resource="#MarvelComics">
      Publisher:
      <span property="publisherName">Marvel Comics</span>
    </div>
  </div>
</section>
```

Figure 8. RDFa Lite markup added to HTML5 content.

While the RDFa Lite syntax includes the basic attributes necessary to embed semantic markup, it is also fully compatible with the complete RDFa Core. In the above example (Figure 8), the `property` attribute is used to express relationships between resources, or otherwise to indicate the predicate in an RDF triple. The `property` attribute requires a URI, Internationalized Resource Identifier (IRI), or a Compact URI (CURIE), which is an abbreviated URI such as "#me". The embedded RDF triples can then be extracted from the content (Figure 9) by an RDFa Processor.

```
@prefix cbo: <http://comicmeta.org/cbo/> .
<#XMen_1963>
       a cbo:Series ;
       cbo:seriesTitle "The X-Men" ;
       cbo:publisher <#MarvelComics> .
<#MarvelComics>
       a cbo:Publisher ;
       cbo:publisherName "Marvel Comics" .
```

Figure 9. RDF triples extracted from RDFa markup.

Despite the relatively small set of properties that compose the RDFa Lite syntax, the syntax appears to be somewhat confusing and difficult for webmasters to successfully implement, and requires prerequisite knowledge of existing Web vocabularies and how to properly use them (Pohorec et al., 2013). Additionally, schema.org identifies the complexity of the syntax as a reason behind the project's decision to focus on Microdata, despite RDFa's extensibility and expressiveness (Schema.org, n.d.). In an attempt to address some of the complexities and usability concerns associated with the implementation of RDFa by webmasters, content authors, and publishers, this study explores the use of metadata application profiles (AP) as a potential solution. Each profile is composed of a subset of elements from the Ontology, and selected properties from established Web vocabularies, presenting solutions for publishing Linked Data in HTML/RDFa designed to meet the functional requirements of select user groups.

2.2.3 Metadata Vocabularies

A metadata vocabulary provides a common set of elements, which data publishers can use to describe information. A metadata element set or schema, is a model for organizing information and may also define rules or standards for the description of resources. Metadata, or data that describes data, serves a number of functions, including resource discovery, organization of electronic resources, interoperability, digital identification, and archiving and preservation (NISO, 2004). In addition to the primary functions of metadata, Gilliland describes a common set of types, or categories, of metadata: (a) administrative, (b) descriptive, (c) preservation, (d) technical, and (e) use (2008). The terms "vocabulary" and "ontology" are often used interchangeably regarding Web vocabularies (W3C, 2013a), although an ontology may define more complex relationships between concepts and other metadata elements, such as classes and properties.

2.2.3.1 Ontologies

An ontology is defined by Thomas Gruber as an "explicit specification of a conceptualization" (1995, p. 908). In this context, a conceptualization represents the objects, concepts, and entities within a domain, and an abstract view of the relationships that exist between them (Gruber, 1995). Gruber notes that formal ontologies are designed, and describes a set of design criteria for the development of ontologies, which includes: clarity, coherence, extendibility, minimal encoding bias, and a minimal ontological commitment (1995). Of note in this list is the minimal ontological commitment, and Gruber states "An ontology should make as few claims as possible about the world being modeled" in order to better facilitate knowledge sharing (1995, p. 910).

2.2.3.2 Resource Description Framework Schema (RDFS)

RDFS is a data-modeling vocabulary that adds additional mechanisms to the RDF model for describing the relationship between groups of resources (W3C, 2014b), and can be used to build metadata vocabularies. This is accomplished through the declaration of *classes*, *subclasses*, and *properties*, which are used to describe a concept or provide refinement to existing components of other vocabularies without having to redefine them (W3C, 2014b). This methodology is particularly useful when implementing general RDFS vocabularies, like Dublin Core (DC) or the schema.org Web vocabulary, where additional refinement is needed or more complex relationships between concepts and entities need defined to better represent the data of a specific knowledge domain. While RDFS provides a basic means for describing connections between resources, the Web Ontology Language (OWL) provides extended capabilities for richer vocabulary definition, the modeling of complex relationships between concepts, and enhanced semantic reasoning.

2.2.3.3 Web Ontology Language (OWL)

The Web Ontology Language (OWL) is a Semantic Web standard that expresses formally defined meaning through the declaration of *classes*, *properties*, and *instances* (W3C, 2012a). OWL also provides enhanced reasoning capabilities through the expression of *domains*, *ranges*, and *cardinality*, as well as the ability to make inferences about *relationships* that exist within data (Segaran, Taylor, & Evans, 2009; W3C, 2012a). The W3C states that there is no clear distinction between the terms "ontologies" and "vocabularies" regarding the Semantic Web, and ontologies can express concepts and relationships about a domain to varying degrees of complexity or simplicity, depending on the needs of the application (2013a). The benefit of using an ontology as a structure for defining a metadata vocabulary is that the formal definition of terms and the taxonomy developed through the relationship of classes and properties represents a more precise way of communicating meaning and understanding between systems (Segaran et al., 2009). A popular tool for authoring OWL ontologies is Protégé Desktop, an open-source ontology editor developed by the Stanford Center for Biomedical Informatics Research. In addition to the desktop version of the software, WebProtégé offers a collaborative ontology development environment for the Web that fully supports OWL 2. OWL 2 expands the OWL language via enhanced data types and ranges, qualified cardinality restrictions, asymmetric, reflexive, and disjoint properties, and improved annotations (W3C, 2012a), among other features.

2.2.4 Usability of Metadata Vocabularies

Because metadata vocabularies represent various degrees of complexity, larger vocabularies can present many usability issues for users. Jakob Nielsen describes the concept of usability, or the user-friendliness of a service, system, or product, as having multiple components, and identifies five traditional attributes: (a) learnability, (b) efficiency, (c) memorability, (d) errors, and (e) satisfaction (1993, p. 26). With regard to metadata vocabularies—where their purpose and primary value is to describe and exchange data—usability may refer to the ability of a potential user to learn and understand that vocabulary, efficiently apply the vocabulary in their work, and successfully apply those terms without error to achieve the expected results.

In a study of the ResearchCyc ontology, Conesa, Storery, and Sugumaran, identified a classification of usability problems associated with the usage of a large, upper ontology describing general concepts and having a wide area of potential applicability (2010). This classification of problematic areas included: (a) understandability, (b) learnability, and (c) operability (Conesa et al., 2010). The authors make a number of suggestions for improving upper ontologies, including the production of better documentation and accompanying visualizations (Conesa et al., 2010). Some of these usability issues identified by Conesa et al., such as understandability, are mitigated significantly by the implementation of domain specific vocabularies, where the terms being identified are recognizable to users familiar with that domain. In a discussion of relating system terminology and vocabulary to user expectations, Nielsen remarks “one should take care not to use words in nonstandard meanings” (1993, p. 123), or that are otherwise unfamiliar. While Nielsen’s argument and advice in this section of *Usability Engineering* has been formulated to address user interface design, much of the author’s rationale for “speaking the users’ language” (1993), can also be extended to the construction of vocabularies and ontologies. However, even large domain vocabularies that serve a particular user community, when accompanied by extensive documentation, can create similar usability issues representing usage barriers for users looking to implement that vocabulary

One approach to improving the usability of metadata vocabularies is the application profile (AP). A metadata AP is a composite set of metadata elements from one or more metadata schemas that expresses the functional requirements of an application or community, while remaining compatible with the base vocabularies from which those profiles are derived (Duval, Hodgins, Sutton, & Weibel, 2002). The concept of APs has been applied successfully to a variety of general vocabularies, particularly Dublin Core (DC), which provides the base schema for the DC-Library Application Profile (DC-Lib). DC-Lib is a compound set of elements that clarifies the usage of DC metadata in libraries (DCMI, 2004). This study applies the AP concept to the study's ontology, and divides its components into a subset of recommend elements for participants within the comic book domain, specifically publishers, retailers, collectors, libraries, and researchers. The intention of these profiles is to guide the publication of structured data by providing recommendations, examples, and other components, including elements from existing Web vocabularies.

2.3 Comic Books and Linked Data

A variety of metadata vocabularies and standards exist for gallery, library, archive, and museum (GLAM) communities, and can be used to successfully describe a variety of resources. These vocabularies include VRA Core for the description of works of visual culture, Encoded Archival Description (EAD) for describing finding aids, and Categories for the Description of Works of Art (CDWA), among many others. Additionally, many research efforts are working towards publishing data described using these standards as Linked Open Data (LOD), as well as exploring metadata quality, enhancement, and interoperability within these communities. Comic books and comic book collections embody many of the bibliographic, visual, and archival object properties of the GLAM community, while sharing many similar areas of focus, concern, and levels of resource description. While many of these vocabularies can be used to generally describe a comic book or related item, no specific domain metadata standard is used for the description of comic books and comic book collections, and no comics-specific Web vocabulary is available for defining relationships between the concepts and entities in this domain, or interlinking resources between various systems.

Comic book data exists in a variety of information systems, databases, and indexes. These systems include many library resources, such as catalogs, finding aids, and datasets in the Linked Data Cloud. Some of these Linked Data resources include WorldCat, which publishes data about many graphic novels and manga found in libraries (worldcat.org), and the Virtual International Authority File (VIAF), a service that provides subject headings and authority data that includes information regarding many comic book creators and publishers (viaf.org). Other datasets, like DBpedia and Freebase contain details about many of the series, storylines, and characters that—when linked to other resources—are capable of enriching existing comic book records (dbpedia.org; freebase.com).

In the preceding examples, the entities described were local, but by replacing the URI identifying those resource, it is possible to link that information to a wider data environment by using identifiers from linked datasets (highlighted below in Figure 10).

```
@prefix cbo: <http://comicmeta.org/cbo/> .
<#XMen_1963>
       a cbo:Series ;
```

```
        cbo:publisher <http://viaf.org/viaf/152451185>.
<http://viaf.org/viaf/152451185>
        a cbo:Publisher ;
        cbo:publisherName "Marvel Comics" .
```

Figure 10. A Linked Data example using VIAF.

This process not only has the potential to introduce enhanced, accessible methods of bibliographic control to the comic book domain, it also has the potential to bring richer, more accurate descriptions to existing databases and catalog records by incorporating data from the wider comics community.

2.3.1 Manga Metadata Framework

A form of comics art and publication that has made great progress regarding Linked Data is that of Japanese manga. Recognizing that manga, similar to the American comic book, has not always been included in library collections and that much of the authoritative data describing the material is not complete, He, Mihara, Nagamori, and Sugimoto used existing Linked Open Data (LOD) resources to identify works of manga in library catalogs (2013). In their study, the researchers demonstrated a process of identifying Functional Requirements for Bibliographic Description (FRBR) work entities in the Kyoto International Manga Museum catalog, by using DBpedia as the necessary authority data (He et al., 2013). Accompanying this effort, the Metadata Framework for Manga explores three aspects of the medium: (a) bibliographic description, (b) structural description, and (c) intellectual entities (Morozumi, Nomura, Nagamori, & Sugimoto, 2009). The framework uses FRBR and TV-Anytime as upper-models, and similar to the Comic Book Markup Language (CBML), focuses on the importance of the structural description of comics content (Morozui et al., 2009). Manga shares many visual similarities to that of the American and European comic book, and appears in these regions as either a translated form of manga, or may be reconfigured as a comic book (O'Nale, 2010). An example of this reconfiguration is the series *Akira*, which was originally published as a manga in 6 volumes by Kodansha, and then as 38 individual comic book issues by Marvel Comics (comics.org/series/3636; see also comics.org/series/15569 for additional example from Dark Horse Comics).

2.3.2 Schema.org and Other Projects

Other efforts to incorporate the world of comic books and graphic novels into the Linked Data ecosystem include a "Comics and Periodical Schemas" extension for schema.org (W3C, 2014d). The proposal includes the addition of schema:PeriodicalSeries, schema:PeriodicalIssue, schema:ComicIssue, and schema:GraphicNovel to the vocabulary, along with related properties for each class (W3C, 2014d). In September 2014, the work of the W3C Schema Bib Extend Community Group was incorporated into schema.org, and included the addition of many of the periodical components listed above, as well as schema:hasPart and schema:isPartOf for describing sequential relationships (w3.org/community/schemabibex).

Both Marvel and DC Comics publish detailed story, event, and character information in Wikis, and Marvel Comics makes additional data available through an API (developer.marvel.com). Notable projects that have utilized the Marvel data service, include the Ultimate 75th an interactive visualization of over 70 years' worth of data (ultimate75th.com), and a graph database of the Marvel Universe produced by Pete Olson, which connected the various relationships between characters in the publisher's stories and illustrated them as a social graph using Gephi, an interactive graph visualization utility (marvelentertainment.tumblr.com/post/30536120271).

Chapter III

Methodology

3.1 Overview

This project presents a case study for the development of a domain ontology for comic books and comic book collections, titled the Comic Book Ontology (CBO). Additionally, the study produced a series of metadata application profiles (AP) that each represent a subset of the ontology and properties from other Web vocabularies intended for use by a select group of participants within the domain. This group of participants included publishers, retailers, collectors, libraries, and researchers. The intention of the CBO Ontology is to function as a tool for linking information about comic books, graphic novels, and other related materials between various systems maintained by these participants and others using Semantic Web technologies. While there are a multitude of options available for the publication of Linked Data in the Semantic Web, this study explores publishing structured data inline by adding semantic annotations to existing HTML content using RDFa. In an attempt to identify the anticipated goals and functional requirements of each user group in the potential implementation of the vocabulary, preliminary user research methods and techniques were used in a review of existing information systems and an analysis of their content. The results were summarized in user personas documents that collected research from multiple phases of the project, and were used to guide the alignment of the Ontology with the needs identified and affinities discovered, as well as to provide focus and direction for the design of the resulting APs. This cluster of properties shared between participants within the same group, and the requirements identified, represents the focus of each profile. Accompanying each AP is a series of solutions and examples in

HTML/RDFa. RDF triples extracted from this content were used to test the viability of each profile and the Ontology itself in describing Web content and linking that content to other resources, especially existing Linked Open Data (LOD) resources containing information about comic books, creators, characters, and collections.

3.2 Research Design

An ontology is an "explicit specification of a conceptualization," and as a conceptualization is an "abstract, simplified view of the world that we wish to represent for some purpose" (Gruber, 1995, p. 908). While this view of the world can be both general and widely applicable, such as the view presented by the Friend of a Friend (FOAF) vocabulary, the perspective of the world offered by an ontology can also be much narrower and more specific. A domain ontology that focuses on the entities and resources in a specific knowledge domain, or area of practice, can provide the level of detail and granularity required by domain experts to share knowledge, map existing data, and exchange information. Additionally, in the larger Semantic Web environment, these detailed descriptions offered by specialist user communities can enrich existing descriptions of the same resource in other systems, even those maintained by large organizations or knowledge institutions. While domain ontologies are useful for capturing and exchanging knowledge, their development requires an understanding of the various agents, concepts, objects, and relationships that exist within the domain. Thus, a case study approach was determined to be the ideal research methodology used in the development of a domain ontology for comic books and comic book collections. "Case studies are analyses of persons, events, decisions, periods, projects, policies, institutions, or other systems that are studied holistically by one or more methods. The case that is the *subject* of the inquiry will be an instance of a class of phenomena that provides an analytical frame—an *object*—within which the study is conducted and which the case illuminates and explicates" (Thomas, 2011). The case study presented was divided into phases: (a) Phase I which built the models and structures, and focused on data conversion, and (b) Phase II which then explored data publication on the open Web.

This research study attempted to conduct a thorough analysis of the domain by capturing a broad and comprehensive view of the world of comic books. In order to obtain an informed view and develop a data model that was adaptable to the many participants within the domain—as well as the multitude of perspectives regarding a comic book object—the study at Phase I began with the first set of tasks in which a comprehensive review was conducted of reference resources, existing projects, and example materials. Secondly, after reviewing the collected resources and materials, a domain model was designed and a series of schemata were developed that reflected this research: (a) an XML schema was built to maintain compatibility with existing systems and function as a map between data; and (b) an OWL ontology was built to explore publishing that data in the Semantic Web using the same, interoperable set of vocabulary terms. Third, a Pilot Study was then conducted to test the viability of these structures in converting data from CSV to XML, then to RDF/XML. An XSLT stylesheet was used to aid the conversion process, replacing common values (e.g., publisher names, languages, etc.) with compatible Linked Open Data (LOD) URIs. Technologies and methodologies used in the production of these structures and in the conversion workflow are discussed within the following sections (see 3.2.4).

Once an alpha version of the Comic Book Ontology (CBO) was published, attention was then placed on using the vocabulary to publish structured data on the Web, recognizing that many of the rich information resources regarding comic books and graphic novels had existing and stable Web content, indexes, and databases available. Phase II of the study then focused on the possibility of applying the vocabulary within the content of these existing systems. In order to ensure the CBO Ontology was capable of meeting the needs and functional requirements identified in this larger Web environment—and assess how this content was actually being published—the study implemented preliminary user research techniques to collect data. After selecting a set of agents from the domain, a review of existing information systems and an analysis of their content was conducted. Major findings were summarized in user persona documents, and each persona was used to guide the alignment of the Ontology's data model with the results discovered and in the design of a series of user-focused profiles. In an effort to improve the usability of the Ontology through the reduction of the vocabulary to a manageable, condensed set of elements, a series of metadata application profiles (AP) were built to address the specific goals and functional requirements identified for each user group. The ability of these profiles and their markup examples to successfully publish Linked Data in HTML using RDFa, from which meaningful, valid RDF could be extracted was then evaluated. The remaining sections of this chapter present the details of the major steps (see 3.4 and 3.5).

3.2.1 Information and Reference Resources for Ontology Development

A number of print information resources and reference materials are available regarding the topic of comic books and graphic novels. Many are accounts of the industry's history, some are price guides and handbooks for hobbyists, and others provide critical analysis and review of both the content and culture of the medium. While a quick scan of the wealth and variety of library holdings regarding this topic would hopefully strengthen the study's argument that the comic book is a complex object, worthy of speculation and research from many domains of study—this project focused primarily on prominent and accessible materials that described the comic book as an object (not a specific series or story), its production, consumption, and those that were commonly used as reference materials in the hobby of comic book collecting. Selected resources used as reference materials during the course of this study and in the development of the Ontology, included works by Scott McCloud, Robert M. Overstreet, and *The Power of Comics: History, Form & Culture*.

Understanding Comics: The Invisible Art by Scott McCloud, published in 1993, is a highly esteemed overview of the comic medium's history, terminology, visual vocabulary, and style. It is one of the first academic studies of the art form of comics, preceded only by *Comics and Sequential Art* by Will Eisner. At the time of its original publication, *Understanding Comics* was particularly groundbreaking because it also presented its content in comic book format. McCloud has since released a set of follow-up works: *Reinventing Comics*, which provides an overview the medium and business of comics and provides insight into its digitization, and *Making Comics*, which functions as a handbook for the production of comic books, manga, and graphic novels. This study used the McCloud trilogy of comics reference and study materials to confirm concepts, establish arguments for the direction of certain aspects of the Ontology's resulting data models, and the most recent entry, *Making Comics*, provided valuable insight in determining contributor roles, various forms and stages of artwork production, as well as the formal differences between comic books, manga, and graphic novels.

The Overstreet Comic Book Price Guide is an annual publication produced by Robert M. Overstreet and published by Gemstone Publications. The price guide contains the estimated values for individual comic books in a range of physical conditions from Near Mint (NM-), which is highly collectible, to Good (GD), meaning that the book displays significant wear but is still readable, or a *reader* copy. These values are estimates and are determined using data and input provided by a group of advisors who also produce an individual "Market Report" each year. While other resources, such as eBay, are often used to determine market value it is still common for books being offered for sale to be given a price of *book* or *guide*, meaning the value is taken directly from a price guide. In addition to pricing, the Overstreet guide also contains information about the content of individual issues such as key character, object, or team appearances, credits of note such as the first or last issue of a particular creator's run, or any other piece of information that may affect the value or significance of a given comic book issue or part of a series. Accompanying the popular price guide, Overstreet also publishes *The Overstreet Guide to Collecting Comics* and *The Overstreet Guide to Collecting Comic & Animation Art*. Both resources provide a comprehensive overview in grading, preserving, and assessing both comic books and original comic book artwork. This study made extensive use of the guide's grading definitions and glossaries in the development of the ontology specification, and the presentation of the guide's listing data forms the initial starting point and benchmark for the Ontology's metamodel. Additional terms and concepts identified in these resources, particularly significant types of archival material, also provided reference for the Ontology's Collection Model (see 4.1.2.5) and intangible concepts that related to archival materials, quality, and condition.

The Power of Comics: History, Form, and Culture by Randy Duncan and Matthew J. Smith, is a comprehensive overview of the comic book medium and a textbook for comics studies and scholarship. This study primarily referenced chapters regarding comic book history, the industry, and the book's overall discussion of the medium and genre itself. The participants selected for the study's system review and content analysis are derived from Duncan and Smith's communication model of comic books (2013). While the study's primary focus is on the comic book object, especially as a bibliographic resource, the insight provided by the work regarding other aspects and perspectives of comics research and scholarship was invaluable, especially for developing an adaptable and extensible model that could eventually be useful in capturing and linking the knowledge from these various user communities.

The above list of information and reference resources represents the primary set of works that were used to shape the ontology and its various components over the course of the study. These resources were also used to validate results discovered in the system review and content analysis that followed in later phases. However, this is not a definitive list of resources used nor is it meant to represent a comprehensive list of available reference materials on the topic of comic books. One additional reference series of note is Marvel Comic's *Official Index to the Marvel Universe*. The 2010 edition of this series collects the contents of a publication of the same title which provides a comprehensive index for each of the individual issues in many of the publisher's popular comic book titles (e.g., Avengers, Spider-Man, Wolverine, etc.). This content includes both a synopsis and notes regarding each issue, in addition to detailed information about the story of each work. These indexes were used to measure the depth of description possible for a given comic book issue, especially with regard to its content, and were used to model other potential access points, such as the comic book characters, objects, locations, and events important to each work.

3.2.2 Gaining Practical Indexing Experience

Understanding how data about comic books and graphic novels is described on the Web requires having practical experience in publishing or creating that data. In order to gain this experience, the researcher volunteered as an indexer for the Grand Comics Database (GCD) during the course of this study. The GCD is a non-profit, international project focused on collecting data about the world's comic books (comics.org). While there are a number of Web databases about comic books and graphic novels that accept submissions, user contributed data, or indexes, the GCD was selected because of the quality of its data and membership. The information in the GCD is multi-lingual, detailed, and all user submitted content is approved by a team of editors. Additionally, the project publishes a formatting specification and implements a mentoring program in which new indexers are paired with a guide in order to receive feedback as they make their first contributions. After accumulating the required number of Indexing Measurement Points (IMPs), indexers are then permitted to submit to the general queue of pending changes where their contributions are approved by the editing team. Changes to the project's policies, indexing specification and guidelines are voted upon by members of the community who have achieved a predetermined amount of IMPs. This model in which: (a) new contributors are paired with a mentor; (b) receive feedback as they submit data; and (c) have those contributions approved by an editor, is an excellent and viable method for maintaining quality and control over user submitted content. Users maintain joint ownership of their contributions with the GCD, and the organization makes all data available under the Creative Commons Attribution 3.0 Unported license (GCD, 2012a).

The GCD is one of many similar efforts in which a dedicated community of users contributes a significant amount of research, time, and effort to the documentation and description of a specialized area of knowledge. As demonstrated by the success of MusicBrainz, a "community-maintained open source encyclopedia of music information" (MusicBrainz, n.d., para. 1), this community created data is valuable to the larger ecosystem of data regarding a specific topic or knowledge domain. The British Broadcasting Company (BBC) is one of many organizations that use data from the MusicBrainz project to enrich their own records (www.bbc.co.uk/music/brainz). Projects that make their data available via open licensing policies (and publish it as Linked Open Data), are subject to receiving many of the same benefits available to libraries and other knowledge institutions that publish data using the same licensing practices, most notably increasing the visibility and usage of those efforts by enabling that data to be remixed, reused, and extended by others (Baker et al., 2011). Ultimately this philosophy of repurposing and reuse is what guided the study to disregard building yet another comic book database or web application to test the Ontology, and instead focus on exploring how that vocabulary can be used with content from existing projects and systems.

3.2.3 Acquisition of Example Materials

In an effort to obtain an objective and well-informed view of comic book objects during the study, the researcher sought to obtain materials that were representative of specific concepts and relationships being explored, especially those that would need addressed by the Ontology. Some of these topics included: adaptations, reprints, aggregates, translations, and artwork. There was no specific approach applied to the acquisition of these materials or criteria used except that they be cost-effective and illustrative of the topic being explored. Three primary examples were assembled over the course of the study and were used as the framework for conducting queries during the system review and analysis, as well as building examples of using the Ontology. Many of the examples embody multiple concepts, in addition to the primary focus they represent. Illustrated examples are available in a companion Web repository (sean.petiya.com/thesis).

3.2.3.1 Example A: X-Men Days of Future Past (Story)

The first example begins with *The Uncanny X-Men in "Days of Future Past"* a trade paperback containing reprinted material from *X-Men (1963)* #141 and *Uncanny X-Men (1963)* #142. This set of collected issues represents the popular "Days of Future Past" story arc by Chris Claremont and John Byrne, which has been adapted into both a movie and cartoon. Found at a New Dimension Comics in Tarentum, Pennsylvania in July 2013, this item represents the start and much of the inspiration behind the study. This particular collection and its parts were selected because they represent a number of interesting relationships between, and occurrences within, comics publications: (a) a story arc that takes place across multiple issues or titles; (b) a change of series title; (c) multiple representations of the same series name; (d) multiple adaptions of the story and artwork; (e) multiple aggregated works and collected editions; and (f) relationships to other media. Both of the individual issues reprinted in this edition were acquired in December 2013 at the Steel City Con in Monroeville, Pennsylvania. Additional collected editions containing the issues were acquired from online retailers. These included *The Uncanny X-Men Omnibus Volume 2*, a large collection of work from this particular comic book series, and *X-Men: Days of Future Past*, a hardcover collection containing both issues as well as related content from other titles, notably parts of the "Days of Future Present" storyline. The final item in this set includes an entry from the World of Reading series acquired at a Toys "R" Us, titled *Days of Future Past*, a children's reader which adapts the story presented in the comics.

3.2.3.2 Example B: Spider-Man (Copy)

The second example begins with *L'Étonnant Spider-Man (1993)* #5, published by Bandes Dessinées Fantastiques. This item was discovered at Liberty Comics in Liberty, Ohio in May 2014 and completed the search for a translation example. This particular issue is a French Canadian translation of *The Amazing Spider-Man (1963)* #302, translated by Julie Maltais. The English version of this comic was then purchased the following June at the Akron-Canton Comic Con in Akron, Ohio. Additionally, a collected edition containing this issue, *The Amazing Spider-Man Omnibus*, was acquired from Amazing Comics & Cards in Long Beach, California via eBay. The final piece in this example set includes a certified copy obtained from an eBay auction in August 2014. The copy is guaranteed by the Comics Guaranty Company (CGC), with its condition graded at a 9.4 or Near Mint (NM), meaning nearly perfect. At the time it was acquired, it represented one of 62 other 9.4 copies of *The Amazing Spider-Man (1963)* #302 according to the CGC's census data, a population which at the time included a total of 314 graded copies of this particular issue overall (cgccomics.com).

3.2.3.3 Example C: Nick Fury (Artwork)

The third, and final, example used in this study includes the original inked artwork of page 8 in *Nick Fury, Agent of S.H.I.E.L.D. (1989)* #17 by inker Romeo Tanghal over pencils by Herb Trimpe. Both the artwork and the issue it appeared in were acquired from a dealer at the Steel City Con in August 2014.

3.2.3.4 Other Examples

A number of other examples were used for reference in this study, of note are a copy of *The Marvel Fumetti Book* #1 found at But-Con in Butler, Pennsylvania in February 2014, and a copy of *Strange Tales (1951)* #167 signed by Jim Steranko at All-Americon in Warren, Ohio in July 2014. Additional handbooks and indexes mentioned were purchased at Spectrum Comics in Boardman, Ohio. To provide reference for the relationship between American comic books and translated volumes of Japanese manga, *Hellsing (2003)* published by Dark Horse Manga and *Ghost in the Shell (1995)* published by Dark Horse Comics, are also used by the study.

While an overview of how these examples were acquired as well as where and when may seem overly detailed, if not excessive, it was written with the hopes of demonstrating how a comic book collection is assembled over time, from a variety of sources, and that a collection in its entirety may include a variety of formats and materials, including comic books, trade paperbacks, graphic novels, manga, sketches, and artwork, amongst other items. Additionally, these examples are intended to illustrate the various bibliographic relationships that exist between comic books and their various manifestations, as well as the complex roles and responsibilities involved in the production and publication of a comic work. These example sets of materials serve as a foundation and reference point for the rest of the study, and the entities and relationships present in their description serve as the framework for designing the CBO Ontology's domain model.

3.2.4 Pilot Study

3.2.4.1 Designing a Domain Model

As evidenced by the examples in the preceding section, any given comic book may be part of one or multiple whole-part, work-to-work, and part-to-part relationships existing between various material types, including trade paperbacks, graphic novels, and other collected editions. This is true not just for the sequential relationship between individual issues and the series or volumes of which they are a member, but each issue itself may be embodied in a number of manifestations including variants, reprints, or formats, especially digital. And of course, translations, where those expressions are typically parts of a unique series with a different publisher than the original material. Of particular importance to this domain is the ability to describe individual copies, or items, with significance placed on physical condition. In order to effectively express these various levels of description inherent to comic books and related materials, the Functional Requirements for Bibliographic Records (FRBR) conceptual model was selected as an upper-model.

The FRBR conceptual model defines four entities in its first group, or Group 1: (a) *Work*, "a distinct intellectual or artistic creation"; (b) *Expression*, "the intellectual or artistic realization of a work […]"; (c) *Manifestation*, "the physical embodiment of an expression of a work"; and (d) *Item*, "a single exemplar of a manifestation" (IFLA, 1998, pp. 16-23). This entity-relationship model is often referred to as WEMI, and provides a vocabulary for identifying bibliographic resources at various levels of description and abstraction, with *Item* being the most concrete as it represents a specific object. This study aligns its domain model of comic books with WEMI (Figure 11), focusing on an individual comic book issue at the Work-level.

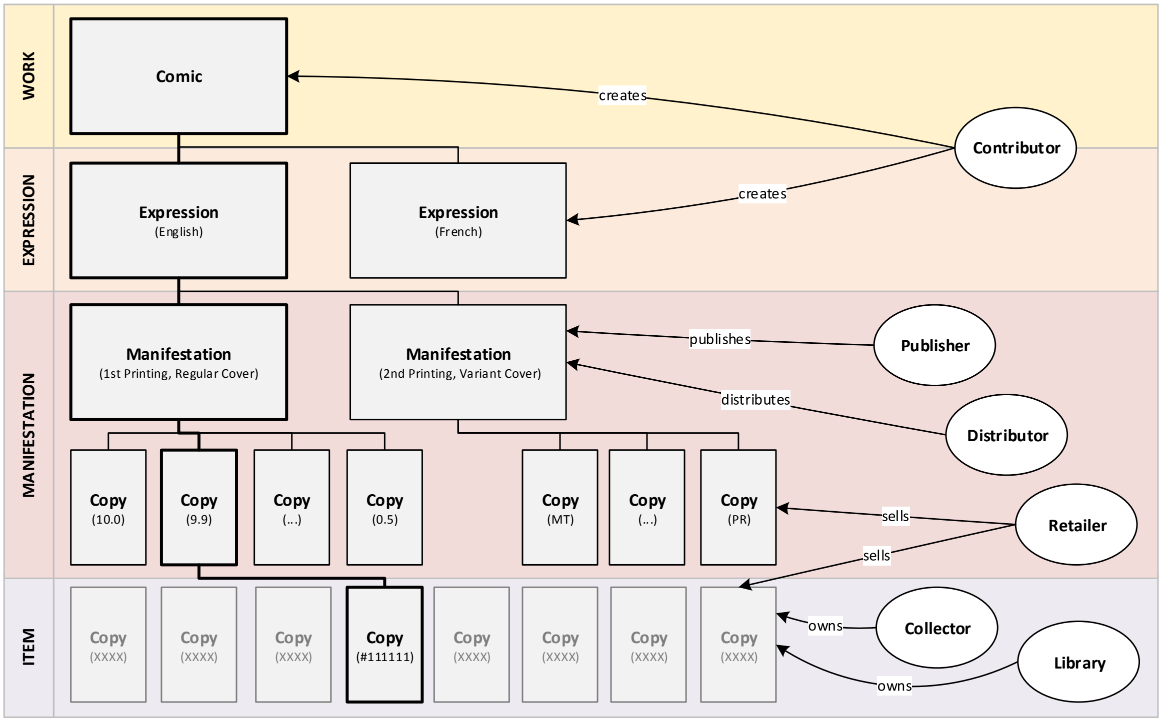

Figure 11. A domain model for comic books based on FRBR.

In this model, Example B's *The Amazing Spider-Man (1963)* #302 is a *Work*, its English and French Canadian translations are *Expressions*, with its format, printing, and edition representing various *Manifestations*. Reference to a copy of this issue in a particular grade or physical condition may still be somewhat abstract, especially when listing estimated values. This reality is illustrated by the second layer of manifestation in the domain model, and addressed in the Ontology's Item Model (see 4.1.2.4). Finally, the *specific*, certified copy owned by the researcher represents an *Item*.

The model presented above is not the first FRBR-based model of comic books. William T. Fee previously identified the relationship between Machine Readable Cataloging (MARC), FRBR, and the potential for enhancing the findability of comic book materials in the library (2008). In Fee's model, the author aligns the comic series with FRBR's Expression, and defines a Manifestation as a *mini-run*, or where "a particular scripter, artist and cover artist worked on all issues" (Fee, 2008, p. 184). This study does not address MARC cataloging or the classification of comic books and graphic novels in a traditional library setting, and the researcher suggests the work of librarians William T. Fee, David S. Serchay, and Randall Scott for reference and insight regarding comics librarianship.

While FRBR is utilized as an upper-model, this study limits its focus and discussion to the alignment of the central object in its domain model with WEMI entities. FRBR Group 2 and Group 3 entities are present in the data and supported by the resulting ontology, but their explicit relationship to FRBR Group 1 (WEMI) is not discussed. It should also be noted that the comic book domain is both collaborative and participatory, meaning that responsibilities are often overlapping. A creator may often assume the role of a publisher, a collector can assume the role of a creator, and so forth. Roles are not mutually exclusive, and any agent may have multiple responsibilities. This reality is also indicated by the multitude of complex relationships present in the derivative works important to comics culture, especially fan-produced content. However, before representing all of these connections and relationships, the specific comic book being described must first be identified.

3.2.4.2 Developing an XML Schema

Data about comic books and comic book collections is stored in a variety of systems, both on the Web and in mobile and desktop applications. Collectors may also use spreadsheet software to catalog their collections. Many of these systems allow users to export or save that content as Comma Separated-Value (CSV) data. Recognizing that this format of data is very common, an

XML schema was developed to map this existing CSV data to a set of core elements defined by the study.

The element set developed represents the properties necessary to describe a resource at all levels of description identified in the study's domain model for comic books (see 3.2.4.1), and the primary components, concepts, and entities of the metamodel. These core elements include: *publisherName*, *imprintName*, *country*, *seriesTitle*, *seriesYear*, *volumeNumber*, *issueNumber*, *language*, *format*, *edition*, *printing*, *variance*, *condition*, *grade*, and *certNumber*. The XML Schema document was authored using Microsoft Visual Studio, a software development Integrated Development Environment (IDE) (Figure 12).

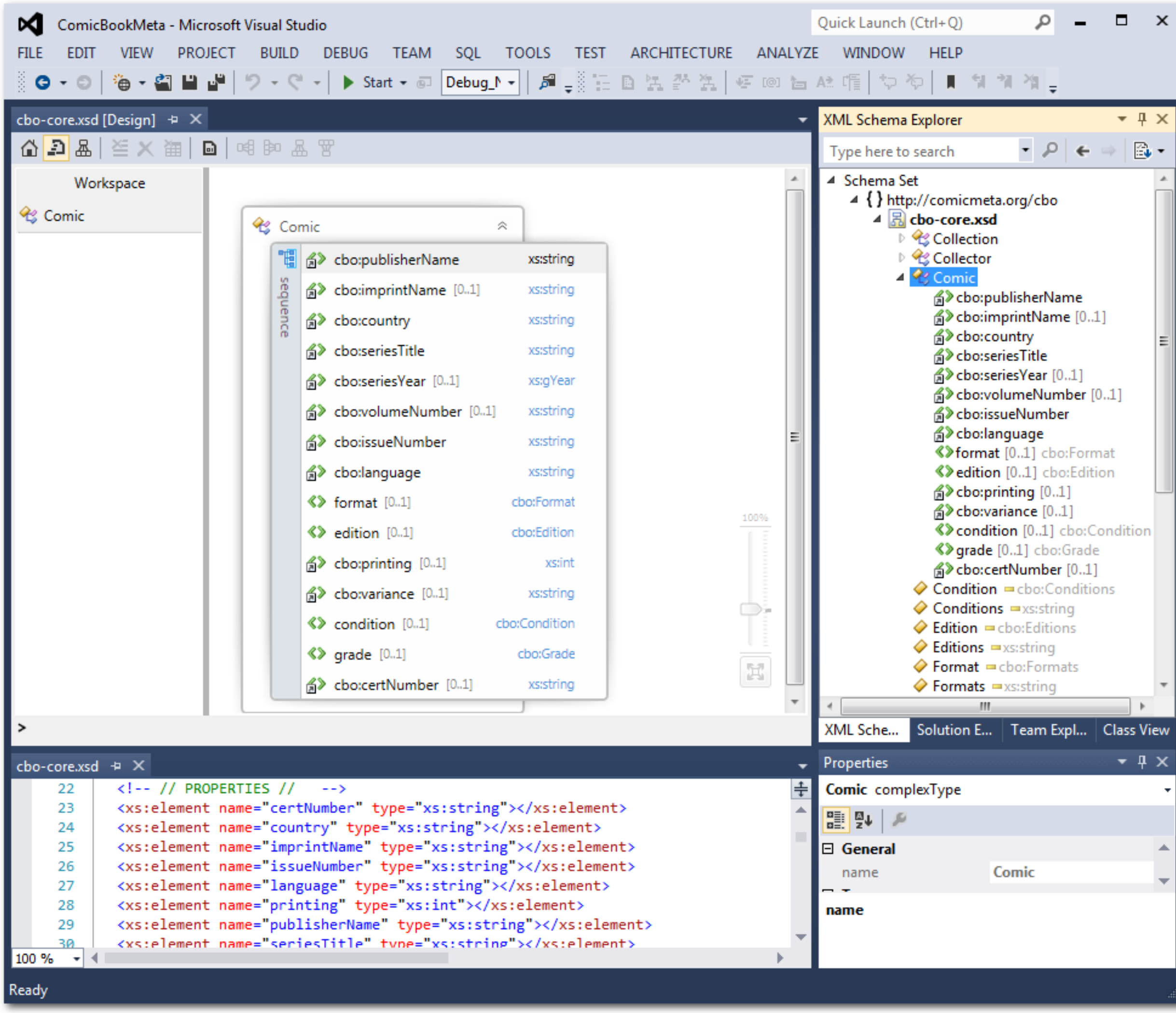

Figure 12. XML Schema editor in Microsoft Visual Studio.

This core schema is a flat, non-hierarchical set of elements that composes a base description, or core profile for a comic book resource and allows for that data to be expressed at multiple levels of description. While the core XML schema was developed to maintain compatibility with existing systems, and to function as a structure for mapping CSV data, it very much mirrors those closed systems and is incompatible with the larger, open Web of Data. In order to extend the ability of the vocabulary to be used in the Semantic Web, an OWL ontology was developed.

3.2.4.3 Developing an OWL Ontology

The Web Ontology Language (OWL) is a Semantic Web standard that defines a number of features for describing ontological structures, including *classes*, *properties*, and *individuals*. OWL documents are primarily exchanged using RDF and can be encoded using a number of syntaxes, including RDF/XML. While an ontological model is similar to an object-oriented data model in that it defines classes and properties, the underlying RDF model differs because it is used to define semantic data which is "focused on the relationships between entities" (Segaran et al., 2009, p. 130). This difference makes ontologies and semantic data "property-oriented" as opposed to object-oriented, where the properties belonging to an entity are determined by its class membership (Segaran et al., 2009). Instead, in a semantic model, the class membership of an entity is determined by the properties used to describe it and a semantic reasoner will infer that a resource belongs to one or more classes based on its relationship to other resources as identified by those properties. This relationship between classes and properties is defined by a *domain*, or the type of class which uses the property, and the *range* or the expected value or data type of the property. In OWL, instances (or members) of these classes, are referred to as individuals. To describe data, OWL provides three types of properties: (a) an owl:DatatypeProperty, with the range of a literal or data value, such as a string of text; (b) an owl:ObjectProperty, with the range of another resource (e.g., an individual); and (c) an owl:AnnotationProperty, which is used to describe or annotate data and is excluded from reasoning. OWL properties are defined independently of classes, and are not required to have a domain or range set. Just as classes can be defined as a *subclass* of another class (a *superclass*), properties can also be defined hierarchically. With this in mind, the Ontology imports both classes and properties from existing Semantic Web vocabularies, or more precisely, Linked Open

Vocabularies (LOV). The Ontology was built using Stanford's Protégé Desktop software (Figure 13), with technical construction guided by *Programming the Semantic Web* by Segaran, Evans, and Taylor and *The Semantic Web for the Working Ontologist* by Allemang and Hendler.

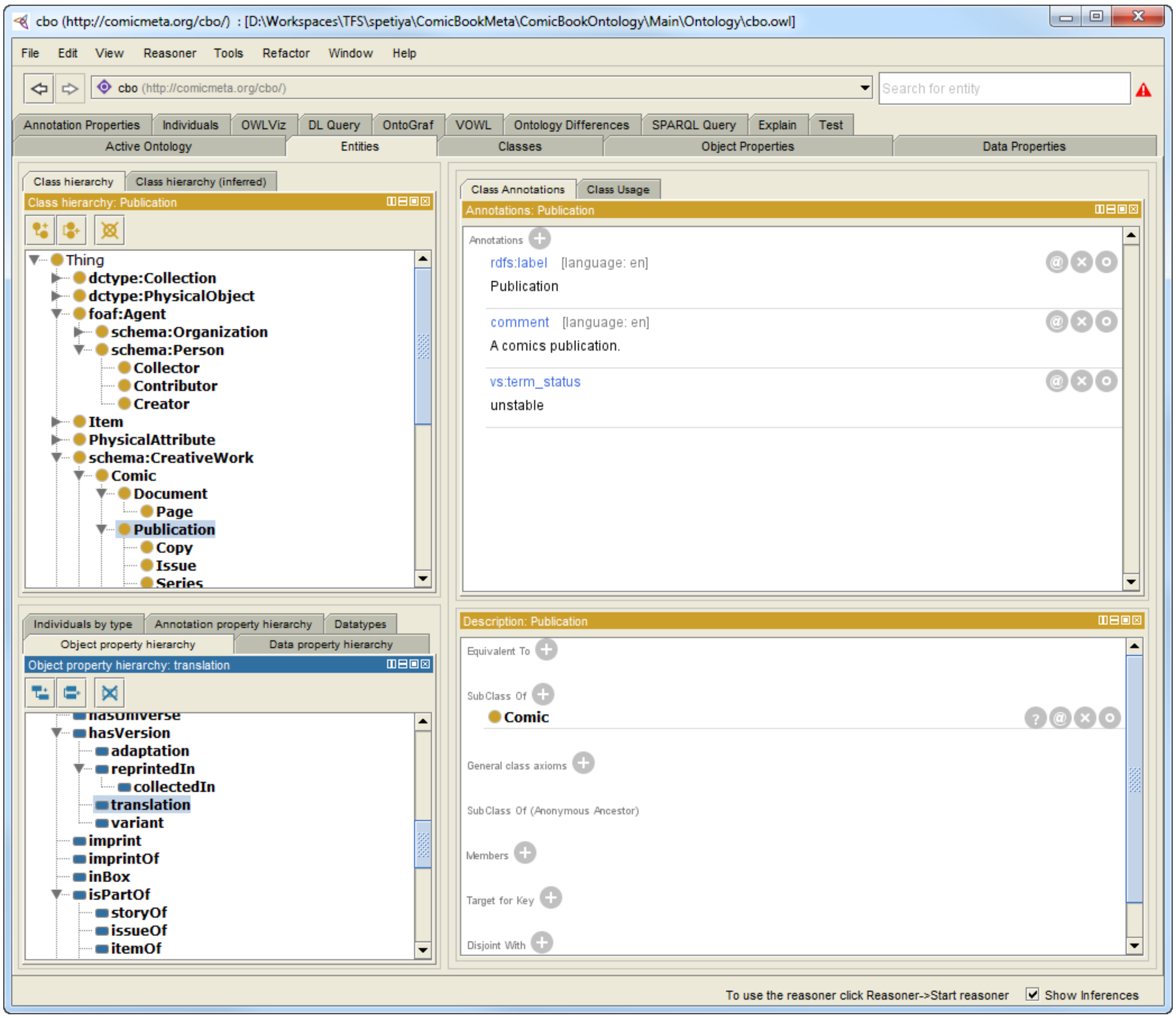

Figure 13. OWL ontology editor in Protégé Desktop.

The Ontology makes extensive use of Dublin Core (DC) and schema.org establishing an upper-model using classes from these vocabularies. This upper-model is intended to define the placement of elements from the Ontology in the larger Semantic Web vocabulary system, for example, a cbo:Comic is defined as a subclass of schema:CreativeWork, and a cbo:Collection is defined as a subclass of dctype:Collection. However, this upper-hierarchy can easily be changed while retaining backwards compatibility with existing data. Even if a cbo:Comic were to be disassociated with schema:CreativeWork (i.e., removed as its subclass) in the Ontology's OWL document, existing data that has applied a schema.org property to a resource, such as schema:about, would still be inferred to be a schema:CreativeWork by a semantic reasoner. This is made possible as a result of the AAA principle ("Anyone can say Anything about Any topic"). While this makes semantic data extensible, it doesn't necessarily mean that data is valid or true. For example, just as one could say an instance of cbo:Comic is also a schema:CreativeWork, bf:Work, or frbr:Work through the usage of any properties belonging to those vocabularies, there is nothing native to the RDF language that prevents a data publisher from stating that a cbo:Comic is also a schema:Volcano. Data validation can be achieved using other methods, such as defining application profiles (AP) or applying constraint languages (see Baker, Coyle, & Petiya, 2014), with validation then conducted by the system consuming that data.

After establishing a set of classes, particularly *Comic*, *Publisher*, *Imprint*, *Series*, *Volume*, *Issue*, and *Copy*, the elements of the XML schema where then added as properties to the Ontology using the above classes as their domain. (Note that the following should be read as: Class (*property*), e.g., Publisher is a "class" which is expressed with a capitalized letter, and *publisherName* is its property): Publisher (*publisherName*), Imprint (*imprintName*), Series (*country, language, seriesTitle, seriesYear, format*), Volume (*volumeNumber*), Issue (*edition, printing, variance*), and Copy (*condition, grade, certNumber*). Where properties were anticipated to be used often, such as *format* and *language*, the terms were redefined (via subclassing) as properties in the Ontology's namespace to limit the number of additional vocabularies a user implementing the CBO Ontology would be required to know or learn. In the final model, presented in the study's Research Findings and Results (see Chapter IV), less emphasis is placed on the distinction between owl:DatatypeProperty and owl:ObjectProperty membership, as a property can belong to both.

3.2.4.4 Metadata Crosswalk to Other Schemas

A metadata crosswalk is defined as "a mapping of the elements, semantics, and syntax from one metadata scheme to those of another" (NISO, 2004, p.11). When a map has been defined between two vocabularies, data can be converted from one specification to another. While this process is useful for exchanging records between systems or converting data, it is not always lossless and some data may lose meaning if improperly mapped, or mapped to a term where the semantics of the destination differ from that of the source element. In an effort to support any existing structured data records, a preliminary crosswalk was made between the metadata elements defined by this study, and existing comic book specifications (Table 1). Mappings defined below are annotated to indicate the type of semantic match represented: (B) indicates a *broad-match*; (N) indicates a *narrow-match*; (C) indicates a *close-match*; and (E) indicates an *exact-match*. Where appropriate, elements in the source schema were mapped to a class in the target schema (the CBO Ontology), especially if that element represented a type of resource (e.g., a panel or balloon).

Table 1. Crosswalk to Existing Comic Book Metadata Formats

CBO	**ACBF**	**CoMet**	**CBML**	**ComicsML**
Class				
Balloon	text-layer (B)		balloon (E)	
Caption	text-layer (B)		caption (E)	

Comic	book-info (B)			comic (E)
Panel			panel (E)	panel (E)
Publisher	publish-info (B)			
Sequence	sequence (E)			panels (N)
ObjectProperty				
balloon	text-layer (B)		balloon (C)	
caption	text-layer (B)		caption (C)	
character		character (E)		character (E)
contributor	activity (C)			
coverPage	coverPage (E)	coverImage (C)		
creator		creator (E)		
editor	activity (B)	editor (E)		
format		format		
genre	genre (E)	genre (E)		
inker		inker (E)		
isVersionOf	source (B)	isVersionOf (E)		
language	languages (C)	language (E)		
letterer	activity (B)	letterer (E)		
page				strip (N)
panel			panel (C)	panel (C)
penciler	activity (B)	penciller (E)		
rating		rating (E)		
role	activity (B)			
sequence				panels (C)
writer	activity (B)	writer (E)		
DatatypeProperty				
date	creation-date (N)	date (E)		
description	annotation (N)	description (E)		description (E)
identifier		identifier (E)		
image		coverImage (N)		
isbn	isbn (E)			
issueNumber		issue (C)		
publicationDate	publisher-date (C)			
publisherName	publisher (C)	publisher (C)		
purchasePrice		price (B)		
seriesTitle	book-title (B)	series (C)		
title		title (E)		title (E)
url				url (E)
volumeNumber		volume (C)		

This map is a direct crosswalk between classes and properties in the CBO Ontology and elements in other schemas that the researcher assumes to be correct, or at least compatible, based on a review of each schema's accompanying documentation. Note that many XML elements from source schemas which are intended to carry literal values are mapped to properties in the Ontology that have a range, or expected value, of a URI. While the data types appear to be mismatched, this mapping would not necessarily produce an error (see 4.1.1). Furthermore, the primary purpose of this exercise was to identify the commonalities between specifications, and if necessary, further iterations of this crosswalk may incorporate more precise mappings. Also important to note is that many of the existing schemas, such as the Comic Book Markup Language (CBML) and ComicsML, are not descriptive metadata languages used for the identification and discovery of resources, but markup languages for encoding the contents of a comic book, specifically the pages, panels, captions, and balloons that compose a visual sequence. Additionally, both remaining formats, the Advanced Comic Book Format (ACBF), and the Comic Book Metadata Format (CoMet), are used to describe a digital object and are intended to identify metadata about a *digital* comic, with CoMet having the widest applicability according to its specification. While this doesn't make any of the above metadata incompatible, it does imply that the semantics of their elements (and specifications) are inherently narrower than those of the Ontology, because the majority imply describing a digital format comic book, and have varying purposes and functions. For example, an encoded XML record containing CBML or ComicsML markup is—in essence—a digital comic.

3.2.5 Data Conversion Workflow

The primary objective of the pilot study was to produce a series of schemata for mapping existing datasets containing comic book data to components of the Ontology. A workflow was developed to achieve this goal and test the viability of using the XML schema as a map between CSV data and RDF/XML data, having previously identified CSV as a common format for comic book records (see 3.2.4.2). This workflow (Figure 14) involves first mapping columns or headers in CSV data to core schema elements, then generating a qualified XML dataset.

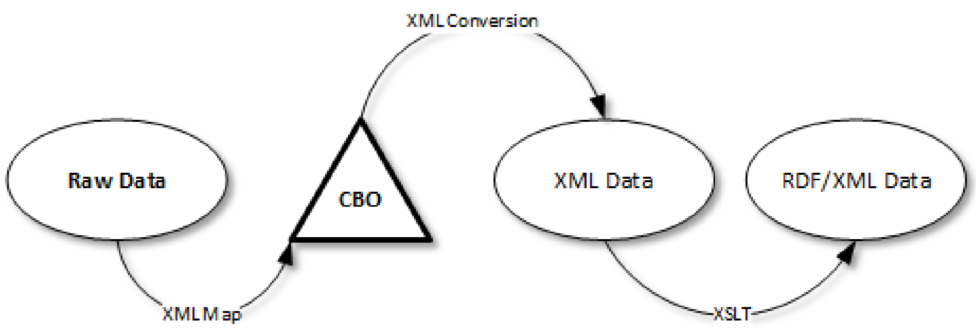

Figure 14. Data conversion workflow.

The initial test of the workflow described above involved using data from Mike's Amazing World of Comics (dcindexes.com). This data is made available in spreadsheet format for reference or personal use, and closely resembles that which a comic book collector might use to catalog their collection. Once the data was cleaned (e.g., quotes removed from titles, NULL, or empty, entries removed, etc.) the workflow described above was then applied to each spreadsheet containing comic book data. Excel was successfully used to map CSV data columns to terms from the study's XML schema, and then generate qualified XML records (Table 2).

Table 2. Results of Sample Dataset Conversion to XML

Dataset Publisher	Version	Total Issues	Resulting Records
Marvel	3/1/2014	40547	35754
DC	3/1/2014	45702	45702
Archie	11/30/2011	10075	10075
Charlton	11/30/2011	6025	6025
Dell	11/30/2011	6503	6503
Gold Key	11/30/2011	5011	5011
Harvey	11/30/2011	6946	6946
		120809	116016

The test resulted in 116,016 XML records produced from the example dataset. Of the datasets converted, only the Marvel set showed a difference between the total issues present in the CSV data and the resulting XML records, possibly due to duplicate records, or entries with empty data as indicated in the “cleaning” process described above. As a comparison to these results, the GCD dataset is multilingual and includes indexes from many countries, for a total of 7,906 publishers, 80,897 series, 1,029,569 issues, and 1,377,012 stories, as of this writing (November 22, 2014). The capability of generating RDF/XML data was then explored.

3.2.5.1 Developing an XSLT Stylesheet

XSLT stylesheets, or Extensible Stylesheet Language Transformations, are used to transform XML documents into other syntaxes, especially HTML. XSLT uses XPath, a query language for XML, to find matching nodes in a target document and then apply one or more templates to the sections of the document that match the pattern. One common usage of these stylesheets is to wrap values from the matched XML nodes in HTML tags for visual presentation on the Web. Similarly, XSLT can be used to produce RDF data from an XML document (Figure 15).

```xml
<xsl:template name="Series" match="cbo:Comic">
  <cbo:series>
    <rdf:Description>
      <xsl:call-template name="Publisher"></xsl:call-template>
      <xsl:apply-templates select="cbo:country"></xsl:apply-templates>
      <xsl:apply-templates select="cbo:seriesTitle"></xsl:apply-templates>
      <xsl:apply-templates select="cbo:seriesYear"></xsl:apply-templates>
      <xsl:call-template name="Volume"></xsl:call-template>
    </rdf:Description>
  </cbo:series>
</xsl:template>
```

Figure 15. An XSLT code snippet.

In the example above, the "Series" template matches all XML nodes of cbo:Comic and applies or calls a series of templates that produce RDF data for matching properties in the Ontology. The study used Microsoft Visual Studio to build the XSLT stylesheet and .NET to transform the data. However, it is important to note that in the current version of .NET (4.5.1), the XSLT engine only supports XSLT 1.0. This is particularly limiting as XSLT 2.0 incorporates significant enhancements to the language. The decision was made by the researcher (for better or worse) to solve all problems using XSLT 1.0 and prevent users who may attempt to use the stylesheet from having to add additional libraries to their project. While this produced some "creative" functions and results, it also hopefully produced some examples with wide compatibility and easily extensible methods. One benefit of the XSLT approach to data conversion is that the stylesheet can be used to transform data using any programming language that supports it (e.g., PHP, Python, etc.). A user can then apply the data conversion workflow to transform data using their preferred methodology and language.

3.2.5.2 Conversion of Data from XML to RDF

While the resulting XML records from the preceding test were used to initially test the ability of the XSLT stylesheet to produce RDF output, a follow-up test was conducted on data exported from the Comic Book Database (comicbookdb.com). This database enables users to create an account and maintain their collection, and most importantly, provides the functionality of exporting that collection as CSV data. Many systems may only provide CSV data exports, especially since using spreadsheet software is a very popular form of cataloging personal comic book collections (see 1.2.3). An example collection of slightly over 50 items was constructed and then modified to include values for country, language, format, series year, and condition. The example below uses *Uncanny X-Men (1963)* #142 from Example A, to demonstrate the first step of the workflow and the resulting XML output (Figure 16).

```
<cbo:Comic>
  <cbo:publisherName>Marvel</cbo:publisherName>
  <cbo:country>US</cbo:country>
  <cbo:seriesTitle>Uncanny X-Men (1963)</cbo:seriesTitle>
  <cbo:seriesYear>1963</cbo:seriesYear>
  <cbo:volumeNumber>1</cbo:volumeNumber>
  <cbo:issueNumber>142</cbo:issueNumber>
  <cbo:language>en</cbo:language>
  <cbo:format>ComicBook</cbo:format>
  <cbo:condition>NM-</cbo:condition>
</cbo:Comic>
```

Figure 16. CSV data converted to XML.

It should be noted that the Comic Book Database includes a number of additional fields in its exported data, but the test of the study's workflow described above was limited initially to only the properties defined in the Core Profile (see 4.2.2). This profile represents the most common elements necessary to identify a specific comic book at all levels of description identified in the domain model (see 3.2.4.1). Applying the XSLT stylesheet to this XML data produces the following RDF output (Figure 17), with common values replaced by Linked Open Data (LOD) URIs (highlighted in the example below).

```xml
<rdf:Description>
  <rdf:type rdf:resource="http://comicmeta.org/cbo/Comic" />
  <cbo:series>
    <rdf:Description>
      <cbo:publisher>
        <rdf:Description rdf:about="http://viaf.org/viaf/152451185">
          <cbo:publisherName>Marvel</cbo:publisherName>
        </rdf:Description>
      </cbo:publisher>
      <cbo:country rdf:resource="http://id.loc.gov/vocabulary/geographicAreas/n-us" />
      <cbo:seriesTitle>Uncanny X-Men (1963)</cbo:seriesTitle>
      <cbo:seriesYear>1963</cbo:seriesYear>
      <cbo:volume>
        <rdf:Description>
          <cbo:volumeNumber>1</cbo:volumeNumber>
          <cbo:issue>
            <rdf:Description>
              <cbo:issueNumber>142</cbo:issueNumber>
              <cbo:format rdf:resource="http://comicmeta.org/vocab/ComicBook" />
              <cbo:copy>
                <rdf:Description>
                  <cbo:condition rdf:resource="http://comicmeta.org/vocab/NM-" />
                </rdf:Description>
              </cbo:copy>
            </rdf:Description>
          </cbo:issue>
        </rdf:Description>
      </cbo:volume>
    </rdf:Description>
  </cbo:series>
</rdf:Description>
```

Figure 17. XML data converted to RDF/XML using XSLT.

In the above example, because viaf:152451185 (Marvel Comics Group) is connected to this comic book series via the cbo:publisher property, a reasoner with the Ontology loaded would infer this resource to be an instance of the cbo:Publisher class. Specific resources are identified using either `rdf:resource` or `rdf:about` (also highlighted in Figure 17). Resources without an identifier are represented using blank nodes (b-nodes), meaning their `rdf:Description` element contains no `rdf:about` attribute. These entities would be assigned an automatically generated identifier when parsed (e.g., genid:A1059666). In this example, the comic book series, volume, issue, and copy have not yet been assigned an identifier. This final conversion workflow and assignment of LOD URIs is aided by a Web-based utility.

3.2.5.3 Data Conversion Utility

The workflow described in the previous sections has been packaged into a simple Web-based utility developed by this study that can be used to convert uploaded data from CSV to XML, or XML to RDF/XML (comicmeta.org/tools/core-convert). Users of the utility are not required to first convert their data to XML, and a method for direct conversion from CSV to RDF/XML is also provided. The replacement of common values with Linked Open Data (LOD) URIs is aided by a series of templates present in the XSLT stylesheet (Figure 18), all of which use a predetermined list of identifiers and common values.

```
<xsl:template name="getPublisherId">
  <xsl:param name="input"/>
  <xsl:variable name="viaf" select="'http://viaf.org/viaf/'"/>
  <xsl:variable name="publisher">
    <xsl:value-of select="translate($input, 'comics group publisher', '')"/>
  </xsl:variable>
  <xsl:choose>
    <xsl:when test="contains('marvel', $publisher)">
      <xsl:value-of select="concat($viaf, '152451185')"/>
    </xsl:when>
     […]
  </xsl:choose>
</xsl:template>
```

Figure 18. Code snippet of XSLT template used for assigning LOD identifiers.

In the XSLT template above, the template “getPublisherId” (a) accepts the publisher’s name as input, then (b) normalizes the string by removing common terms, such as “comics”, “group”, and “publisher”, and finally (c) executes a series of `xsl:choose/xsl:when` statements, which are the XSL version of conditional statements (e.g., if-then, if-then-else, etc.) to find a pre-selected LOD identifier for a specific publisher.

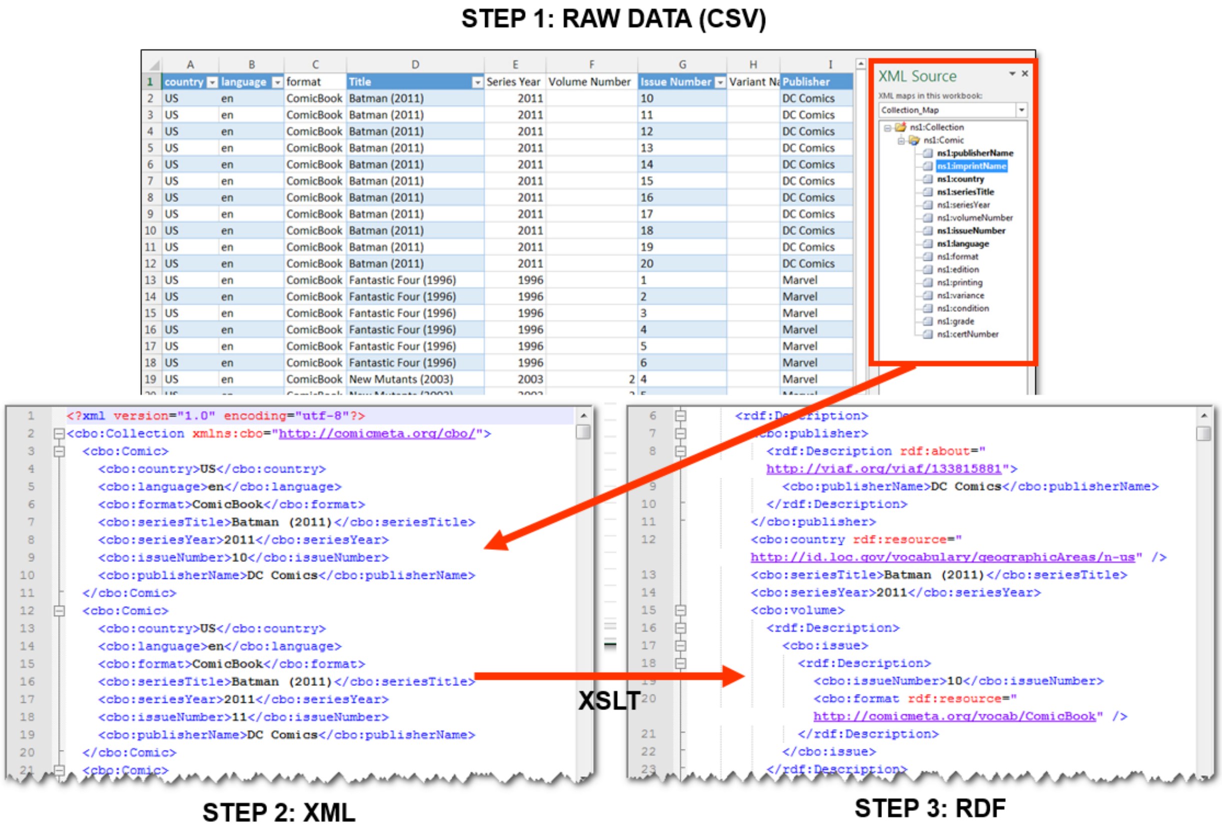

Figure 19. *Diagram of steps in the XML to RDF data conversion process.*

The complete workflow is illustrated in the above diagram (Figure 19) and demonstrates the conversion of raw, CSV data exported from an information system or maintained in a personal catalog (Step 1) to qualified XML (Step 2), and then to RDF/XML (Step 3), with common values replaced by LOD identifiers (see also Petiya, 2014).

3.3 Building a Semantic Web Application

After completing the pilot study, the original final phase of the project intended to build a web application that implemented the CBO Ontology and was capable of demonstrating how data could be exchanged and linked between information systems. However, after having the opportunity to become more familiar with projects like the Grand Comics Database (GCD), the Comic Book Database, and Mike's Amazing World of Comics, it seemed unnecessary to introduce an additional database, application, or index of comic book data when such interesting and notable efforts had already been established. Instead, in Phase II the study shifted focus to these existing information systems and methodologies for publishing Linked Data about comic books and graphic novels on the Web.

The study identified the publication of Linked Data in existing Web content using HTML and RDFa as a potential solution for implementing components of the Ontology within existing data and information systems. This also makes the benefits of Linked Data more accessible to webmasters and content authors, without necessarily requiring the maintenance of an additional database, extensive knowledge of programming, or complete comprehension of Semantic Web technologies and standards. Rather, structured data can be embedded into the markup of existing HTML content and Web applications without making exhaustive changes to backend systems or data stores. Methods for improving the usability of the Ontology and its potential usage were then explored in a review of existing information systems and an analysis of their content.

3.4 Data Collection

3.4.1 Identification of Agents

Agents were first identified using Duncan and Smith's communication model of comic books (2013). The authors identify six components in the model, four of which are enacted by specific participants and include the: (a) *source* (artist, writer, or contributor); (b) *gatekeeper* (editor, publisher); (c) *delivery* (distributor, retailer); and (d) *receivers* (Duncan & Smith, 2013, p. 7). In order to fulfill the objectives of this study, the receivers are assumed to include comic book readers, collectors, researchers, and libraries. This model is limited to the production, distribution, and consumption of a comic work, and diagrams a relatively short life-cycle. However, both the comic book medium and community are participatory. Agents often have multiple and overlapping roles (i.e., a creator can also be a publisher, a collector can also be a creator, etc.), and many works are often reimagined, reprinted, and repurposed by both fans and publishers alike. Fan content, contributions, collaborations, and collections are significant to this domain, the history of the medium, and the longevity of the works. And today, reviews and other editorial content are crucial for maintaining the "buzz" surrounding modern books and stories. However, this study narrows its focus by using Duncan and Smith's model as a framework, and limits the review and resulting profiles to five agents, or user groups. These groups are believed to offer a unique description of the resource in regard to the segment of the life-cycle they are most closely associated with. The following agents were included in the study:

- **Publishers** – Organizations responsible for the publication of comic book materials.
- **Retailers** – Organizations selling comic books and related items.
- **Collectors** – Organizations or individuals that collect comic book materials.
- **Libraries** – Organizations responsible for the maintenance, organization, and retrieval of comic book materials.
- **Researchers** – Individuals interested in the history of comic books.

From this model of comic book communication and the agents identified above, a set of candidate systems for review were then selected based on their perceived relationship to each agent (e.g., publisher and digital library websites, research project indexes, etc.).

3.4.2 Review of Existing Systems

A review of existing systems was conducted in an effort to acquire an understanding of the capabilities and parameters of these applications and their functionality. Emphasis was placed on systems that were fully accessible over the Web, with desktop and mobile applications excluded for the purposes of this review. The review selected systems that appeared to have: (a) quality data, (b) were regularly updated, or (c) had an active community of users maintaining their content. Focus was placed on content that was immediately available without a subscription or purchase. In an analysis of these systems, the review established 22 criteria that measured the available user tasks, additional features, and the structure and markup of the content itself. The criteria (Appendix C.2) selected was believed to have implications in the design and potential application of the CBO Ontology, and was considered relevant in gaining insight into how the related agents publish their content or otherwise make that content available on the Web. The review was limited to 20 systems, four for each of the five user groups (Figure 20).

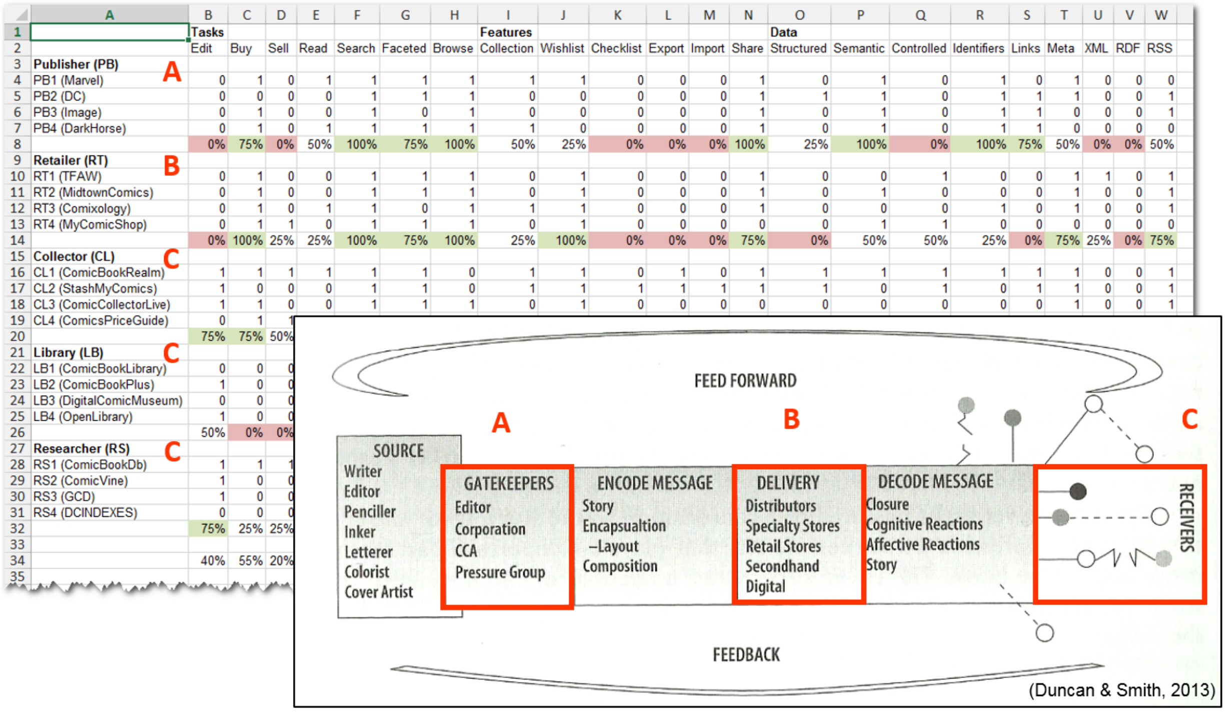

Figure 20. Diagram of the agent and system review selection process.

It would be impossible to include all of the publisher and creator information systems available, as well as all of the research projects or digital libraries containing comic book material. However, the systems selected are fairly representative of the groups to which they belong and exhibit very common information elements, navigation, and content. Additionally, since the framework used for this study is derived from Duncan and Smith's communication model, systems primarily focusing on social or editorial content such as reviews, previews, etc. were excluded. While this dialog between readers and creators is essential to participation in the medium, and important for documenting the life-cycle of these materials—if not extending it—other vocabularies handle blog posts, articles, and reviews quite comprehensively. Future studies may wish to focus on exploring methods for connecting this conversation, or feedback channel, to the graph of comic book data (see 5.4).

In order to conduct the system review, the researcher created an account for each system and searched for each of the listed criteria in the system's interface. Criteria regarding data was uncovered by examining each system's HTML markup, or source code. A spreadsheet was used to track the analysis, and if the system met the criteria being measured, that section was assigned a 1. If it did not exhibit any signs of that criteria, the section was assigned a 0. Each criteria was calculated, and those shared by all systems in the group were determined to have the most significance. The list of systems reviewed and results are presented in Appendix C.

3.4.3 Content Analysis

Once a selection of information systems was made from the review, an analysis of their content was then conducted. A content analysis "reveal[s] patterns and relationships within content and metadata that can be used to better structure, organize, and provide access to that content" (Morville & Rosenfeld, 2007, p. 242). This process of reviewing existing documents and objects, can be an "informal survey or detailed audit," and is described as a "useful tool for learning about the scope and nature of content" by Morville and Rosenfeld (2007, p. 241). In *Information Architecture for the World Wide Web*, the authors use the approach to describe a bottom-up methodology for the development of an information architecture, and recommend gathering a representative sample of content using a "Noah's Ark approach" in which one example of each type of content available is examined (2007, p. 241). For the purposes of this study, the content analysis was limited to Web pages that displayed three specific types of content: (a) details about a single comic book issue; (b) details about a work that was an aggregate of multiple issues; and (c) a list of multiple items. The purpose of this exercise was to discover what details and information each agent published about a comic book, and the focus of the analysis was limited to the visual display of this information in each system. A list of queries was formulated to retrieve pages containing this content from each group of systems (Appendix D.1). Where possible the same query was used for each system belonging to the same group. While Example A was used as the primary search query for Retailers (RT), Collectors (CL), and Researchers (RS), the query was modified for an analysis of content from Publishers (PB) and Libraries (LB). Not all publishers contained data from Example A, and that section of the content analysis required modification. Similarly, the content in Example A (X-Men) is not public domain making it unavailable to many digital comic book libraries, which required the query to be modified and a compatible title common to the majority of systems to be selected (Black Hood).

Isolated areas of information or "chunks" containing relevant data on each Web page were examined. These sections of content were the primary focus based on the understanding that the markup of these areas could potentially carry embedded semantic annotations or structured data. Areas of content that were related, such as series details or a list of characters appearing in the story, were grouped together and assigned a specific area number. Each piece of relevant data within that area was then assigned a numeric identifier. Morville and Rosenfeld note that the content analysis process is "quite unscientific" and recommend starting with a short list of items, such as structural, descriptive, and administrative metadata (2007, p. 243). The study adopted this approach and each one of these *data points* was assigned a color based on the type of metadata it represented (see Appendix D.2). This process was conducted using FastStone Capture, which facilitates capturing and annotating screenshots (Figure 21).

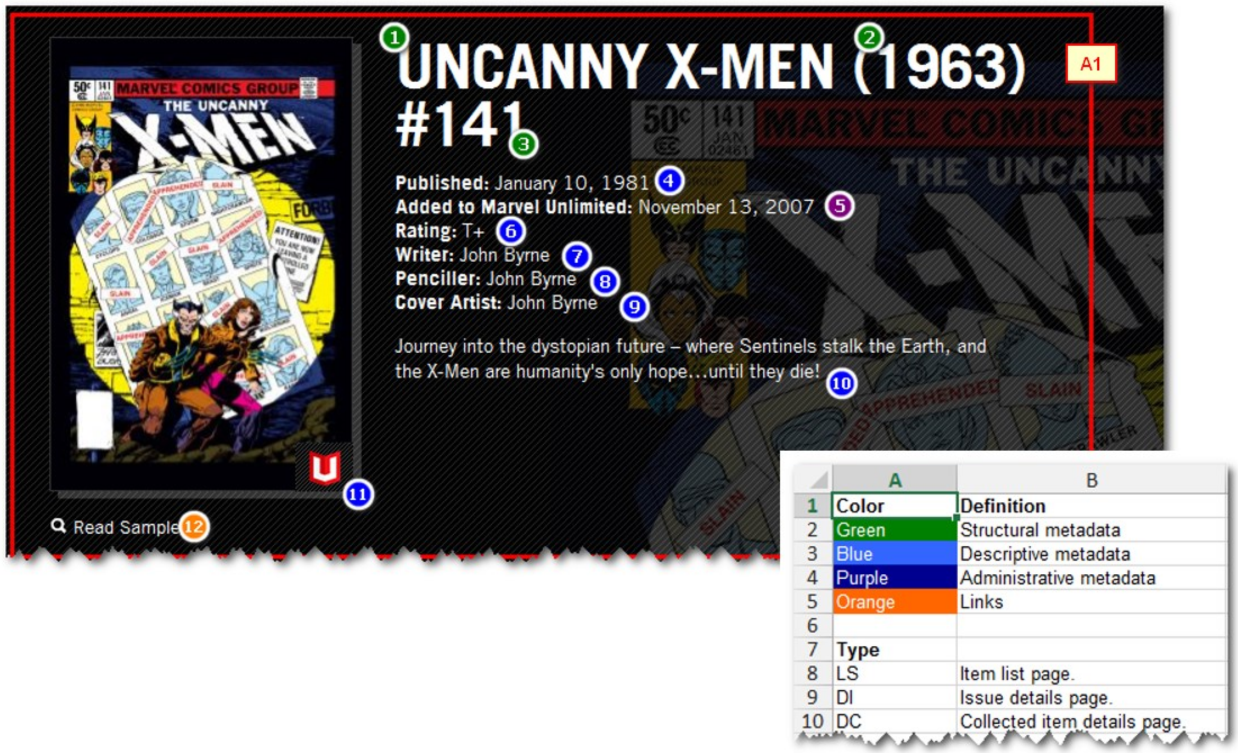

Figure 21. Annotated content analysis of an issue details page.

Using a spreadsheet, each of the annotated data points was assigned a label identifying the information it represented. The results of this process were then indexed for each system by group and type of content (see Appendix D). Morville and Rosenfeld cite the law of diminishing returns in reference to the usefulness of a content analysis (2007), and where the details were overlapping or no longer returning new data points, those areas were ignored in subsequent analysis. This occurred primarily in aggregate works where the relationship to a reprinted work was indicated by only one data point or link, and this type of content was dropped after the PB and RT analysis as those pages were no longer returning unique information.

The index of content items analyzed for each group is presented in Appendix D.3, and the index of all annotated data points and their labels is presented in Appendix D.4-D.8. In addition to the pages of Web content described above, for each system that enabled a user to export data or implemented a controlled term list, the values and fields from these items were also incorporated into the content analysis and taken into consideration during the analysis of the results and alignment with the Ontology (see Appendix F).

3.5 Data Summary and Analysis

All of the data points discovered as a result of the content analysis process were merged by system and imported into separate database tables using Microsoft Access. Using the database software to facilitate the analysis, all tables were joined by group and distinct values were extracted using a series of SQL queries. A list of all data points for each group was then produced (Appendix D.4-D.8). These distinct sets of values were compared after an audit of duplicates, erroneous labels, administrative metadata, and items that were determined to be better addressed by other technologies (e.g., hyperlinks to ecommerce retailers). A master list of values was produced after merging all adjusted lists for each group (Appendix D.9).

A comparison of the data points between all user groups revealed that the shared pattern consisted of: *cover image*, *series title*, *issue number*, *format*, *publication date*, and *page count*. This was significantly different than the original profile and assumed common pattern of data consisting of a: *series*, *volume*, and *issue*. The concept of a publication's volume is used inconsistently within the domain. This is true on the Web, in print, and in the material itself. Further analysis of the adjusted data points revealed that, in addition to the unique requirements of each group (e.g., linking product to retailers, expressing condition and value, etc.), many of these needs and contributions in terms of information overlapped. Data points from each user group were then merged to reveal the similarities between agents. For example, merging values from the PB and RT group displayed that in addition to the shared pattern identified above, these systems also required the ability to display information about the artist and writer of a work. In comparison, an analysis of CL and RS systems revealed that this pair required the ability to describe additional roles, including colorist, editor, inker, and letterer. Content between each pair of groups was also very similar, with the CL and RS systems displaying the most detailed and granular information. LB systems displayed the most unique requirements, such as being able to express provenance and genre. Additionally, a summary of all data points revealed that, prior to adjustment, CL systems had the highest number of data points overall, followed by RS. Overall, depending upon the role of each agent regarding its position in the communication model of comic books, and the life-cycle of the material, the view of the object and needs for describing it change. The anticipated goals and functional requirements identified were summarized in user persona documents (Figure 22), which were used to guide the design of a series of application profiles (AP) that address these major findings and results (see 3.5.2).

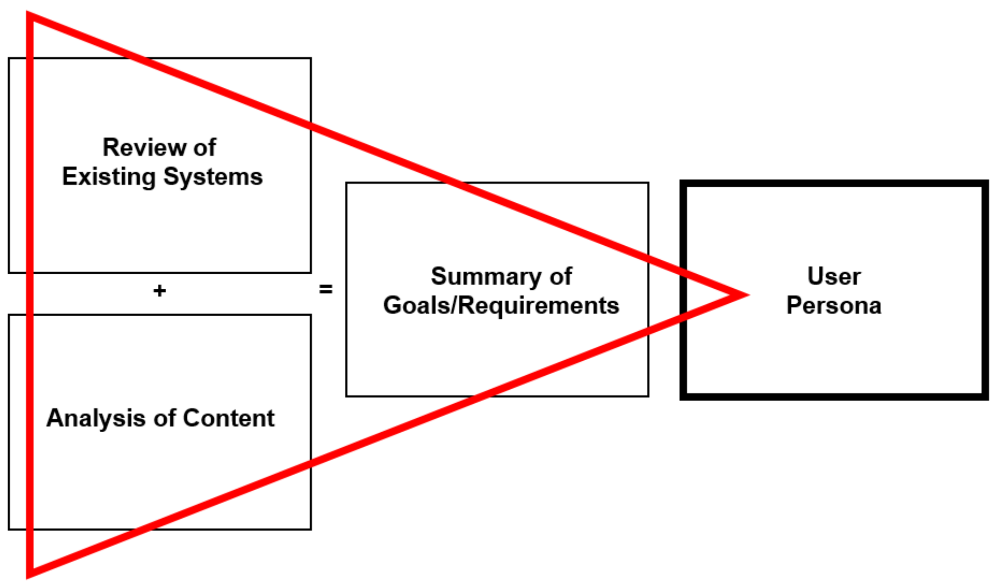

Figure 22. Triangulation of research methods and summary of findings into user persona.

During delivery (PB, RT), the comic book is a product, requiring an on-sale date, price, and retail outlet. In this mode it is also an inventory item often requiring an item number, SKU, and UPC. Once received (LB) it becomes not just a product, publication, or periodical, but also a story, complete with a cast of characters, objects, and locations that form a separate universe of content. As a collection item (CL), emphasis is placed on the quality of the item itself and grade, condition, and ownership become important data points. When studied (RS), emphasis may be placed on the sequences of artwork, or pages of content. These varying dimensions and perspectives are taken into consideration during the alignment of the Ontology with the results of the study and in its final data model.

3.5.1 Ontology Alignment

Using the master list of data points extracted from the content analysis, the Ontology was aligned with these findings (see Appendix F). Each data point was matched to an existing property and if a data point could not be matched, this item was considered as a potential candidate for inclusion in the Ontology. Clusters of data that were better expressed in other vocabularies, such as those describing a product or library holding, were matched to properties in existing Web vocabularies. The best example of this occurs in the ecommerce model inherent to Retailer (RT) and some Collector (CL) systems that offer the ability to sell items. These concepts are thoroughly addressed by GoodRelations (GR) and schema.org. Similarly, data points that were common to other vocabularies or Web content in general, were matched to those well-defined properties (e.g., schema:logo, schema:brand). The exception to this practice, however, occurs when the property is anticipated to be used with great frequency or is part of the Ontology's Core Profile (see 4.2.2). For these instances, the property was either added directly to the Ontology's namespace or included as a subproperty of a term in the source vocabulary (e.g., cbo:seriesTitle is defined as a subproperty of dcterms:title; see also Figure 24).

This alignment process shaped the final data model of the Ontology, the results of which are presented in the study's Research Findings and Results (see Chapter IV). Further analyses of the systems, content, and data of these agents were triangulated into user persona documents that summarize the anticipated needs and functional requirements identified for each user group.

3.5.2 User Persona Documents

Persona creation is defined by Pruitt and Adlin as a process in which data is summarized, clustered, and analyzed to discover themes, the results of which are then used to create outlines or "skeletons" of individual users that can be used for planning, design, and development (2010, p. 156). The authors define skeleton personas as "very brief, typically bulleted, lists of distinguishing data ranges for each subcategory of user," with their purpose being to shift focus to the details summarized in the document (2010, p. 184). After the review of existing systems was completed, and results of the content analysis were clustered and analyzed, major findings were then summarized in individual user persona documents for each group (Figure 23).

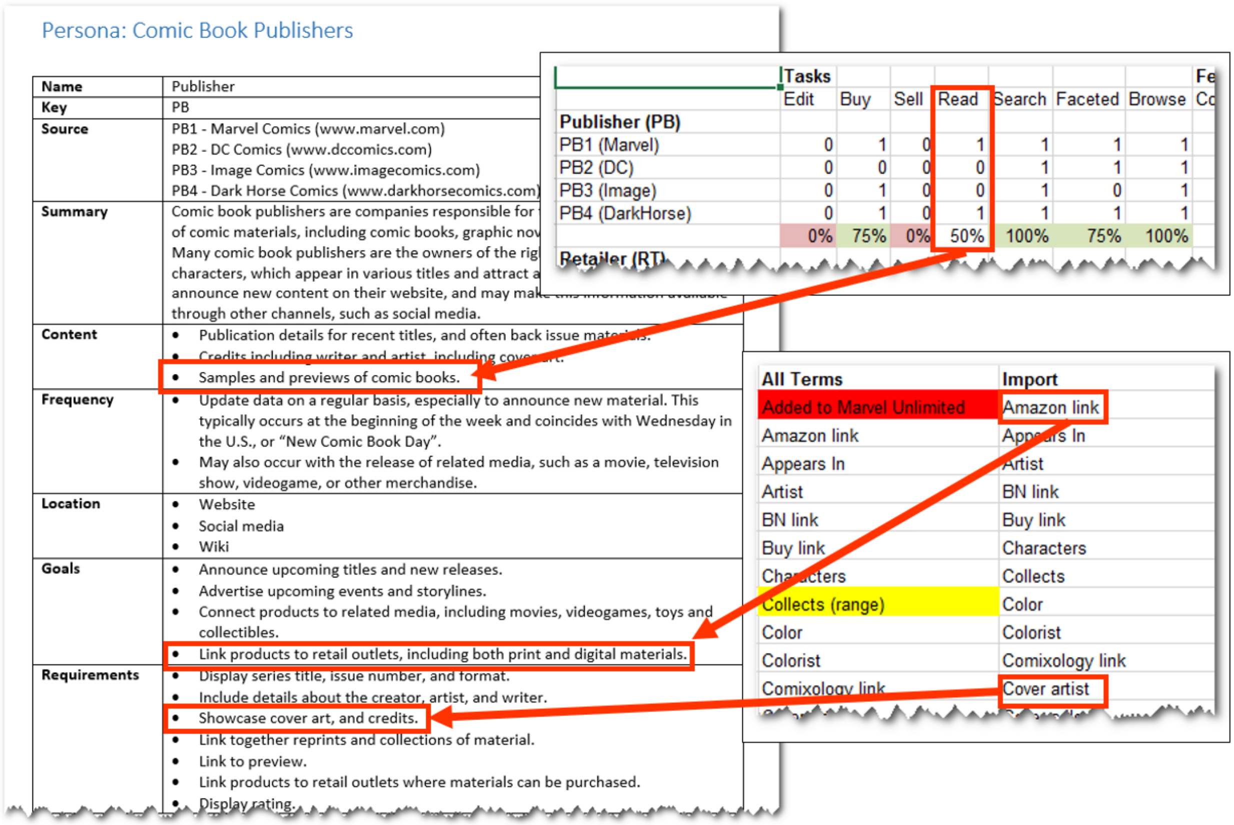

Figure 23. Summary of goals and requirements in a user persona document.

Brown identifies three types of personas: data-driven, institutional, and procedural (2011, p. 37). Additionally, Brown defines three layers of detail: (a) requirements, (b) relationships, and (c) humanization, or "Making 'em Human" (2011, p. 38). This study summarized the results of the case study (i.e., the overview of literature, reference materials, systems, and content) by creating data-driven, skeleton personas from the key data points identified, focusing on the first of Brown's layers. Personas can be iterative or "living" documents, with additional details added following subsequent research. Additionally, they can be divided and further refined into sub-categories of users, such as collectors focusing on graded material, or professionals. The resulting persona documents for this study summarize: (a) who the group is; (b) what content they publish; (c) why they publish that information; (d) when they publish it; and (e) how it gets published. The personas are intentionally simple and provide headings for name, key, source, summary, content, frequency, location, goals, and requirements (see Appendix E). Each document serves as an aggregation of the major findings from each phase of the research process, and were used to provide both focus and a central point of reference during alignment of results with the ontology, and in the design of metadata application profiles (AP) that address the requirements discovered.

Chapter IV

Research Findings and Results

This project presents a case study for the development of a domain ontology for comic books and comic book collections. In its initial phase, the study incorporated a review of information and reference resources, example materials, and existing projects. After a domain model and core metadata schema were built, a pilot study was conducted to test the ability of the structures to produce meaningful RDF from CSV and XML datasets. Once the data conversion workflow was determined to be viable, focus was shifted to publishing RDF data in existing HTML content using RDFa. Potential participants that may choose to implement the vocabulary were identified, and a review of existing Web information systems and content was conducted. This process was used to identify the needs of these user groups in publishing structured data about comic books on the Web, and to ensure the Ontology was capable of meeting their anticipated goals and functional requirements. Results were summarized in user persona documents, and used to guide the design of a series of metadata application profiles (AP) representing subsets of the Ontology and properties from other Web vocabularies intended to meet the requirements of each group. The final result is a flexible and extensible OWL ontology and RDFS metadata vocabulary for describing comic books and comic book collections in the Semantic Web.

4.1 The Comic Book Ontology (CBO)

The Comic Book Ontology (CBO) is an OWL 2 ontology built in Protégé Desktop, and served in a Microsoft Internet Information Services (IIS) hosting environment using custom software built in C# and ASP.NET. The Ontology is hosted in the namespace: http://comicmeta.org/cbo/. Although the host network provides name services (DNS), the domain name comicmeta.org is owned and maintained independently by the researcher. Content negotiation is also handled by the custom software, with dynamic conversion to specific formats provided by the dotNetRDF library (dotnetrdf.org). Finally, documentation is generated using a series of XSLT templates applied to the RDF/XML encoding of the specification. The Ontology itself is composed of a series of classes, properties, and individuals representing various entities, agents, objects, and relationships in the comic book domain.

4.1.1 CBO Properties

The Ontology defines properties of three types: (a) object properties (owl:ObjectProperty), (b) data properties (owl:DatatypeProperty), and (c) annotation properties (owl:AnnotationProperty). Object properties describe the relationship between entities (resources), such as that between a comic book series and its publisher, and data properties express literal values, such as the text of a series title or a volume number. Annotation properties are used to annotate, or document, resources and are excluded from semantic reasoning. Resources in this context are “anything that can be identified with a Universal Resource Identifier (URI)” (Segaran et al., 2009, p. 65). This “anything” is purposefully broad, and the RFC 3986 document which defines the standard states that a URI can be used to identify not just networked resources, but also an “abstract or physical resource” (Berners-Lee et al., 2005, para. 3).

As mentioned in previous chapters, semantic data is property-centric focusing on the relationships between entities, as opposed to the class-centric nature of object-oriented systems (Segaran et al., 2009). Semantic resources are “understood to be members of a class because of their properties” (Segaran et al., 2009, p. 130). This relationship can be indicated by the *domain* (type to which it can be applied) and *range* (expected type) of a property. Thus, where the class membership of an object in a closed system defines the properties that *can be used* to describe it, the class membership of an object in an open system is determined by the properties that *are used* to describe it. This distinction is important because it speaks to the flexibility of semantic data modeling. Just as the properties defined in this Ontology can be used to describe data about comic books within a conceptual model of comics, so can the properties in any other Web vocabulary. This allows a resource identified as a comic work (i.e., a publication, story, or artwork) to also be described as a product, library holding, auction item, gallery piece, subject of an essay, etc. using a conceptual model more appropriate to that context, or perhaps more accurately, one that fully expresses the purposes and objectives of the target application. The result is that a resource identified using a URI can be described by multiple systems, and from a multitude of perspectives, without disrupting the view required or provided by that of another system, fully embodying the AAA, OWA, and NUNA principles of the Semantic Web (see 2.2.1).

With these concepts in mind the Ontology imports a series of properties from existing vocabularies, notably dcterms:hasVersion, dcterms:isVersionOf, dcterms:hasPart, and dcterms:isPartOf. These properties are important for expressing the bibliographic relationships inherent to comic works (e.g., reprints, translates, adapts, etc.), as well as the whole-part, part-to-part, and aggregate relationships between various concepts (i.e., an issue is part of a series, an item is part of a collection, etc.). This collection of properties also has an *inverse*, or opposite, property attached. Inverse properties define a relationship in both directions (W3C, 2012a). The decision was made to subclass the hasPart and isPartOf properties from the DC namespace as opposed to the schema namespace, in order to distinguish the applicability of these properties beyond the description of Web content. However, this decision does not imply that schema:hasPart or schema:isPartOf are incompatible, and just as a resource can be a member of multiple *superclasses*, so can properties belong to multiple *superproperties*. For applications that require these terms to be distinguished by namespace, the necessary maps between qualified properties can be built. This hierarchy can also be used to establish any necessary logical entailments, especially those required to better facilitate search. For example, the Ontology links reprinted work using the property cbo:reprints (or cbo:reprintedIn), which is defined as a *subproperty* of dcterms:isVersionOf. A similar relationship exists between large collections of reprinted material (e.g., trade paperbacks, omnibuses), but is often stated as “collects” instead of reprints. In order to provide the functionality to express this distinction, the Ontology establishes a property called cbo:collects (or cbo:collectedIn), which is defined as a subproperty of cbo:reprints. This hierarchy of properties asserts that a collected work is a reprint, and a reprint is also a version of a work (Figure 24, inferences are highlighted in yellow).

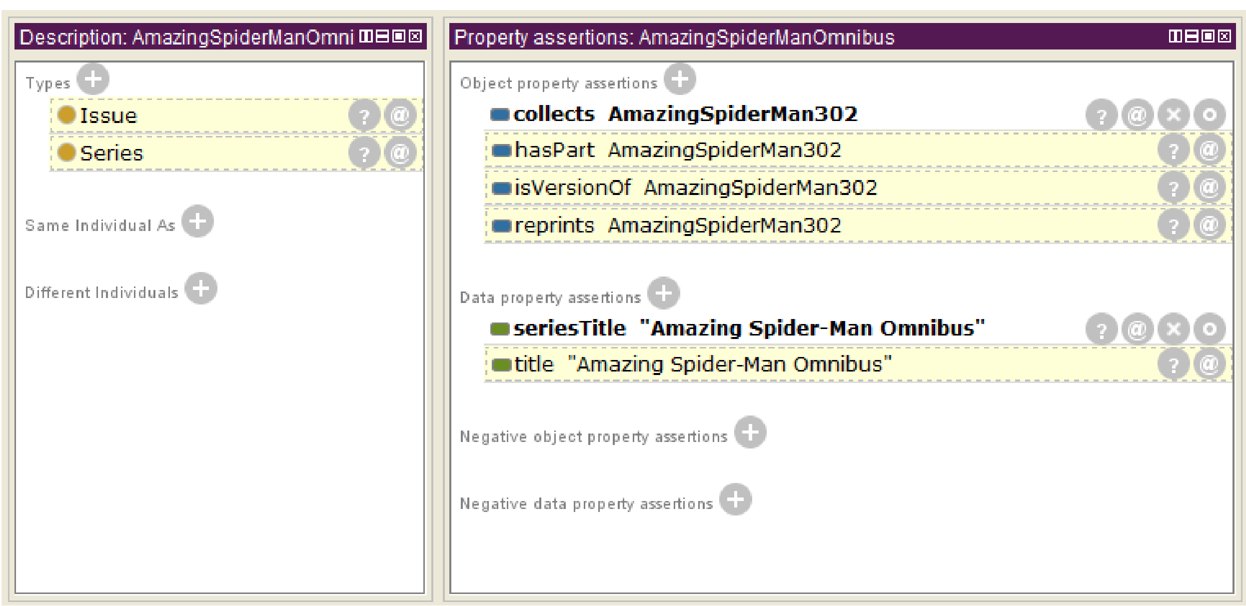

Figure 24. Example of inferences for reprinted/collected works in Protégé.

Users can begin their search using terms at any position in the hierarchy (dcterms:isVersionOf), or make their search more precise using a semantically narrower term (cbo:collects).

While the relationships described between these resources are defined as object properties, in the final specification less emphasis is placed on the distinction between object and data properties regarding their range. An object property used with a literal value would merely be inferred to be both an owl:ObjectProperty and an owl:DatatypeProperty (or owl:AnnotationProperty) by a reasoner. While this clarifies that OWL is not intrinsically a constraint language, it also demonstrates its usefulness on the open Web where it cannot be anticipated that all users will correctly markup their data, or that all users will identify a separate, logical resource for each entity being described. This pragmatic approach is used by schema.org in the description of the vocabulary's data model, which also cites the mantra of "some data is better than none" (2014, para. 2). Validation and the ability to identify values of a specific type can be handled locally by the system, service, or application consuming that data, and specific requirements can be defined in an application profile (AP) or similar technical document.

Connections between works are just some of the relationships requiring description in the comic book domain. Also of importance are the relationships between works and the agents that create, produce, and publish them. With regard to creating and producing a work, the world of comics has a unique vocabulary with terms that include: *colorist*, *inker*, *letterer*, *penciller*, and *plotter*, among other variants of these roles. Additionally, of particular importance to describing the content of a comic work is the *appearance* of a character, team, or object, and the occurrence of an *event*, within its story. And for describing a collection of comics, properties that define the *condition* and *grade* of these items are necessary. The Ontology seeks to provide a complete list of properties necessary to describe these relationships, and anticipates adding additional terms as necessary (see Appendix G). The entities these relationships link together are defined by the Ontology's class hierarchy and resulting data models.

4.1.2 CBO Classes

The Ontology establishes a total of 57 classes, most of which are placed into a hierarchy of superclasses imported from existing Web vocabularies. These imported classes include members of the DCMI Type Vocabulary (prefix: dctype), Friend of a Friend (prefix: foaf), and schema.org (prefix: schema). An entity was added as a class if it was considered to have relevance within the domain, could be described using additional properties and relationships, or was particularly significant to search. The results of this process are presented as a series of data models, including models for: (a) Agent, (b) Work, (c) Universe, (d) Item, and (e) Collection. Corresponding RDF/XML examples are also provided in the following sub-sections.

4.1.2.1 Agent Model

The Ontology's Agent Model (Figure 25) incorporates foaf:Agent as the primary superclass, and introduces schema:Organization and schema:Person as subclasses. Classes are defined for *Contributor, Creator, Collector, Guarantor, Library, Distributor, Museum, Publisher,* and *Imprint*. Data publishers and content authors looking for more descriptive properties regarding these entities can use components of both the FOAF and schema.org vocabularies.

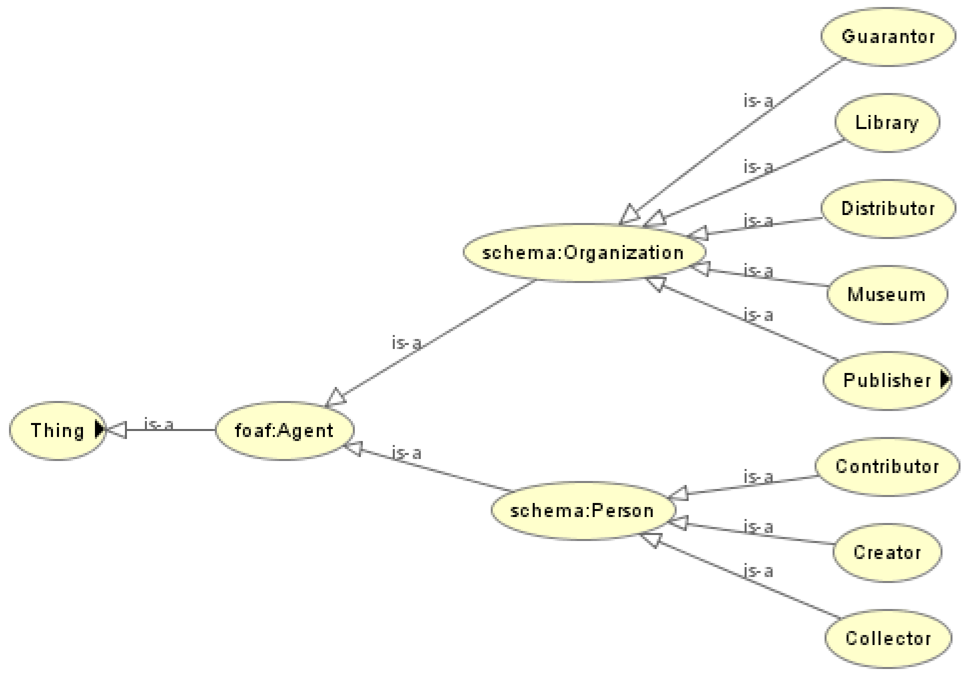

Figure 25. CBO Agent Model.

4.1.2.2 Work Model

The Ontology's Work Model (Figure 26) includes *Comic*, and a series of subclasses named *Publication*, *Document*, *Sequence,* and *Story*. These classes represent the various dimensions, or perspectives of the comic book previously identified: (a) a bibliographic publication; (b) a document containing sections and pages; (c) a visual sequence of pictorial content; and (d) a story. Additionally, this class imposes no restriction on whether "comic" refers to a manga, graphic novel, trade paperback, comic strip, comic artwork, etc. These concepts can be expressed using the cbo:format property, which can be repeated or combined with other properties like cbo:binding, to provide a more precise description (e.g., a hard-cover, black and white graphic novel). The only prerequisite for membership to this class is that the resource is a work of sequential art.

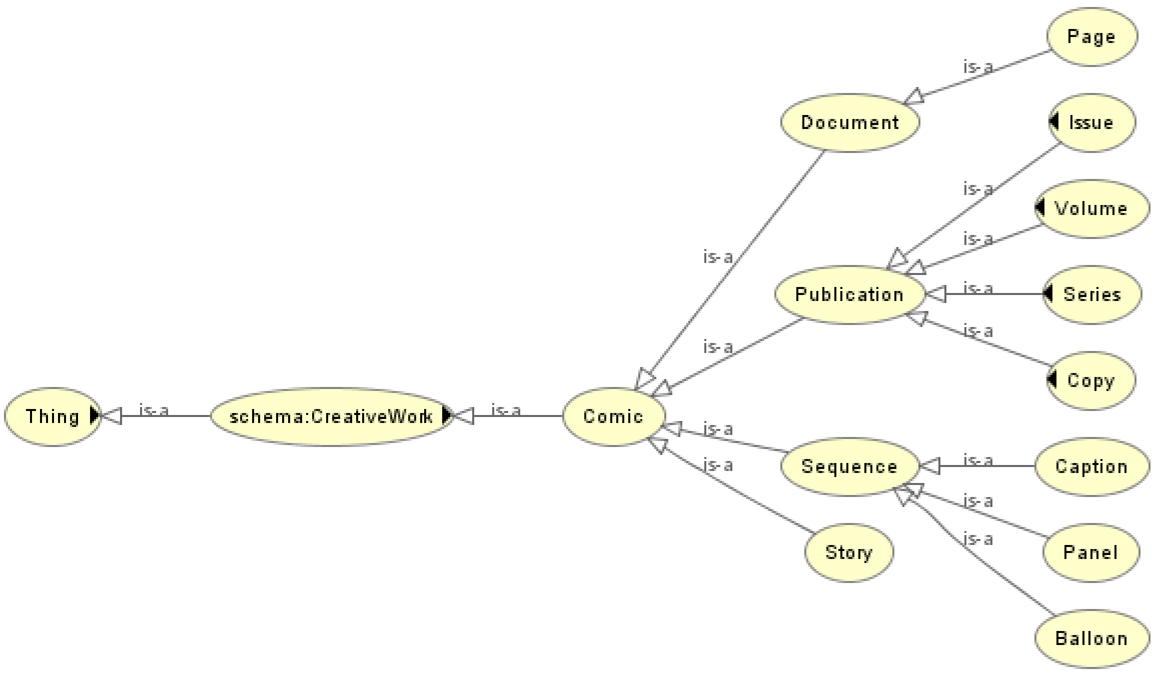

Figure 26. CBO Work Model.

Important to consider here is that the above is not an entity-relationship model. The objects in this diagram do not represent tables in a database, instead they represent ontological classes or more accurately, ontological concepts. A database or application is not required to define an entity for each one of these concepts in order to be compatible with the Ontology. Rather, resources are inferred to belong to one or more of the above classes based on the properties used to describe them. Essentially, they function as containers for describing that particular dimension of a comic work. For example, as previously identified (see 3.5), the concept of *volume* is used inconsistently in the domain. In Example A, the comic book series to which these materials are connected is commonly considered to be volume 1 of *The X-Men*. The *Overstreet Price Guide* refers to this title as "Series 1". The Grand Comics Database (GCD) splits this title into two separate series: *The X-Men (1963)* and *The Uncanny X-Men (1981)*. And in later issues of the series, the publisher identifies the volume as 1 in the indicia, but today refers to it as both

Uncanny X-Men (1963) and *Uncanny X-Men (1963-2011)*. This overlapping description of series and volume also occurs between volume and issue, especially when identifying trade paperbacks and manga. However, the Ontology's Work Model presents a solution to this problem:

```
<rdf:Description>
  <cbo:series>
    <rdf:Description rdf:about="http://www.comics.org/series/2605">
      <cbo:seriesTitle>The Uncanny X-Men</cbo:seriesTitle>
      <cbo:alternativeTitle>Uncanny X-Men (1963)</cbo:alternativeTitle>
      <cbo:alternativeTitle>Uncanny X-Men (1963-2011)</cbo:alternativeTitle>
      <cbo:seriesYear>1963</cbo:seriesYear>
      <cbo:endDate>2011</cbo:endDate>
      <cbo:volumeNumber>1</cbo:volumeNumber>
    </rdf:Description>
  </cbo:series>
</rdf:Description>
```

Figure 27. CBO Work Model RDF/XML example (Publication).

In the above example, the resource is described using both a series year and a volume number, thus a semantic reasoner would infer it to be both a cbo:Series and cbo:Volume. An application parsing this graph could then determine locally how best to consume the data presented and whether or not to treat the resource as one or multiple entities. A similar approach can be taken with the *Story* dimension:

```
<rdf:Description>
  <cbo:series>
    <rdf:Description rdf:about="http://www.comics.org/series/2605">
      <cbo:issue>
        <rdf:Description rdf:about="http://www.comics.org/issue/35133">
          <cbo:issueNumber>142</cbo:issueNumber>
          <cbo:story>
            <rdf:Description rdf:about="http://www.comics.org/issue/35133/#209337">
              <cbo:storyTitle>Mind Out of Time!</cbo:storyTitle>
              <cbo:storyArc
                   rdf:resource="http://dbpedia.org/page/Days_of_Future_Past"/>
            </rdf:Description>
          </cbo:story>
        </rdf:Description>
      </cbo:issue>
    </rdf:Description>
  </cbo:series>
</rdf:Description>
```

Figure 28. CBO Work Model RDF/XML example (Story).

In this example, the story of this particular comic book issue from Example A has been identified using a GCD resource, and the story arc of which it is a part has been identified using a DBPedia identifier. We can also take this example one step further and connect this comic book issue to the wider bibliographic data environment by identifying works that reprint or adapt it:

```
<rdf:Description>
  <cbo:issue>
    <rdf:Description rdf:about="http://www.comics.org/issue/35133">
      <cbo:reprintedIn rdf:resource="http://www.worldcat.org/oclc/44674970"/>
      <cbo:adaptedBy rdf:resource="http://www.worldcat.org/oclc/858126746"/>
    </rdf:Description>
  </cbo:issue>
</rdf:Description>
```

Figure 29. CBO Work Model RDF/XML example (Bibliographic Relationships).

These examples illustrate how both the publication and story dimensions of a comic can be described using the Ontology. Similar approaches can be taken to describe the pages of a comic book, or their *Sequence*. The cbo:Sequence class and cbo:sequence property can theoretically be used to connect the description of a comic's publication, story, etc. on the open Web with encoded markup of its contents (i.e., the panel, word balloons, and captions) contained in an XML document using CBML or ComicsML.

4.1.2.3 Universe Model

As evidenced in the example above, and in the results of this study, it is nearly impossible to separate the comic book from its content. The stories, characters, objects, and locations of its narratives compose a very intricate and detailed world, representing important access points for collocating, retrieving, and discovering materials. The appearance of an object, death of a character, continuation of a storyline, along with many other plot and story elements, are very important components to the search for and enjoyment of comic books. This issue becomes even more pronounced when attempting to uncover the individual issues in a collected edition, determine which works reprint parts of a story arc serialized across multiple titles, or piece together an often very chaotic and elastic narrative. To address these needs, the Ontology separates the “Comic” from the “Comic Universe” in its Universe Model (Figure 30).

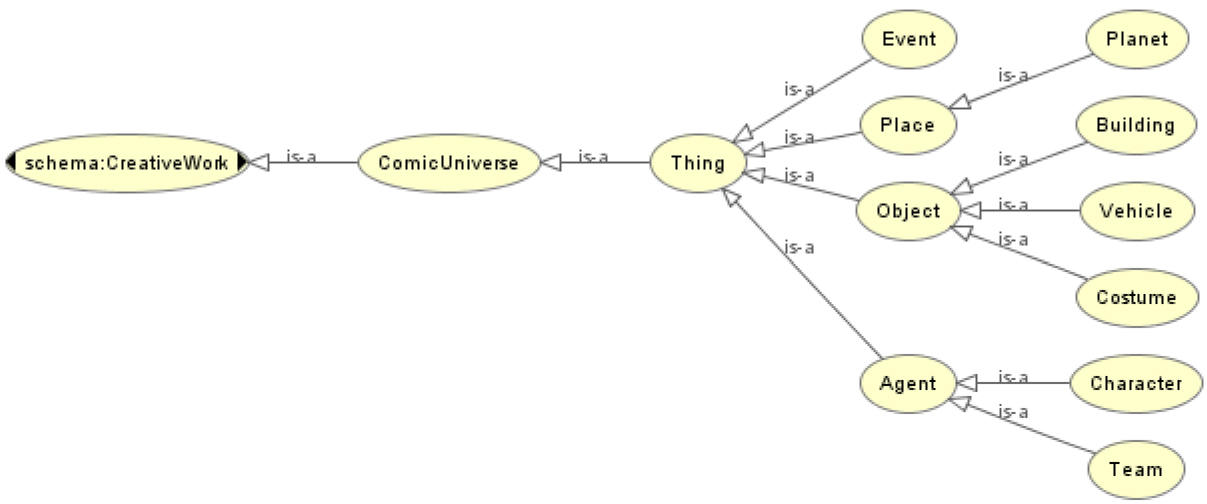

Figure 30. CBO Universe Model.

A cbo:ComicUniverse is a universe of comic book characters, events, objects, and locations created by the various publications, artwork, and other media that reference and compose it, and represents a total summary of this collective, creative endeavor. Like a cbo:Comic, it is a subclass of schema:CreativeWork, but is itself not a work of sequential art and thus not a cbo:Comic. In order to achieve this level of description, the Ontology redefines a new “Thing” within the scope of a “Comic Universe” simply as: “A thing in a comic universe”. This class can then be used as a container for all of the various *things* in that universe. This definition makes no distinction between whether or not the thing is real or fictional, and is not an attempt to completely model a fantasy world. Comic books and graphic novels often contain representations of real people, places, things, and events, and may even be auto-biographical. Pope John Paul II, musician Alice Cooper, and talk-show host David Letterman have all appeared as characters in comic books. This fact does not challenge the nature of their existence, it merely asserts that their likeness, similar to a caricature or avatar, has appeared as a character in a comic book story. Thus, much like an ontology is a "cartoon universe" (Brickley & Miller, 2014, para. 2), the cbo:ComicUniverse class is, quite literally, a cartoon universe inside of a cartoon universe.

```
<rdf:Description rdf:about="http://www.comics.org/issue/35133/#209337">
  <cbo:storyArc rdf:resource="http://dbpedia.org/page/Days_of_Future_Past"/>
  <cbo:storyTitle>Mind Out of Time!</cbo:storyTitle>
  <cbo:character>
    <rdf:Description rdf:about="http://dbpedia.org/page/Wolverine_(comics)">
      <schema:name>Wolverine</schema:name>
      <cbo:universe>
        <rdf:Description
            rdf:about="http://marvel.wikia.com/Earth-811"></rdf:Description>
      </cbo:universe>
    </rdf:Description>
  </cbo:character>
</rdf:Description>
```

Figure 31. CBO Universe Model RDF/XML example.

In the example above, a comic book character from the story in Example A has been identified (Figure 31). Additionally, the *specific* comic book universe to which this character (or version of) belongs has also been identified. This data does not imply that the character is to be precluded from appearing in videogames, movies, or other works, it simply asserts that a representation of this character appears in the story being described. Finally, while the Ontology's Universe Model does not attempt to fully describe a fantasy world, these fictional properties are not without their utility. A character's superpowers, agility, strength, age, etc., can be used by fans, creators, and editors to maintain artistic accuracy and continuity within a series, as well as with multiple titles (especially over many decades).

4.1.2.4 Item Model

The Ontology's Item Model includes two subclasses: *Copy* and *Artwork*. Both are members of the cbo:Item class and the cbo:Comic class. In this domain "copy" may still very much be referring to a manifestation, specifically in a particular grade or condition. Chaining together these properties is how the value of an item is determined. For example, a copy of Example B's *Amazing Spider-Man (1963)* #302 in near mint (NM-) condition, or graded at a 9.2, is listed in the 44th edition of the *Overstreet Comic Book Price Guide* as being worth an estimated $18.00 USD (2014, p. 427). For comparison, a copy in good (GD) condition, or graded at a 2.0, is worth an estimated $2.00 USD (2014, p. 427). These "copies" in the *Overstreet* listing are still referring to the notion of a specific comic book, or more precisely, a manifestation of a manifestation. However, once we have identified a particular copy by serial number, owner, or collection, we can then describe it as a specific, concrete item (Figure 32):

```xml
<rdf:Description>
  <cbo:series>
    <rdf:Description>
      <cbo:seriesTitle>Amazing Spider-Man (1963)</cbo:seriesTitle>
      <cbo:issue>
        <rdf:Description>
          <cbo:issueNumber>302</cbo:issueNumber>
          <cbo:copy>
            <rdf:Description>
              <cbo:certNumber>022672202</cbo:certNumber>
              <cbo:owner rdf:resource="https://sean.petiya.com/#me"/>
            </rdf:Description>
          </cbo:copy>
        </rdf:Description>
      </cbo:issue>
    </rdf:Description>
  </cbo:series>
</rdf:Description>
```

Figure 32. CBO Item Model RDF/XML example.

In this example, the researcher's 9.4 copy of *Amazing Spider-Man (1963)* #302 is identified. This particular copy would be inferred to be a specific *Item* by a semantic reasoner, because the cbo:certNumber and cbo:owner properties both have a range of cbo:Item.

4.1.2.5 Collection Model

The Ontology's Collection model (Figure 33) includes two components: a *Collection* and a *Container*. A cbo:Collection is a subclass of dctype:Collection, and cbo:Container is defined as a subclass of dctype:PhysicalObject.

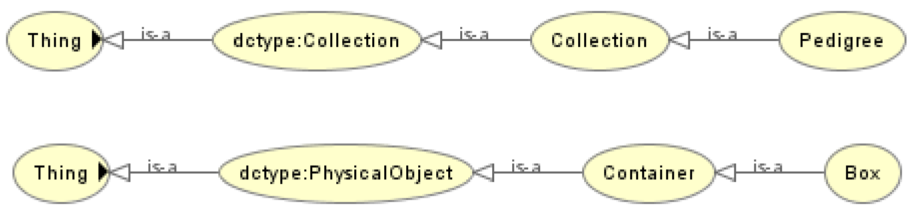

Figure 33. CBO Collection Model.

As discussed in the background of this study (see 1.2.3), items in a comic book collection are often kept in archival boxes specifically designed for comics, commonly referred to as either a *short box* or a *long box*. Collectors may also choose to *bag* and/or *board* their items using archival materials of various quality.

```
<rdf:Description rdf:about="#MyComicBook">
  <cbo:certNumber>022672202</cbo:certNumber>
  <cbo:grade rdf:resource="http://comicmeta.org/vocab/NM-"></cbo:grade>
  <cbo:encased rdf:resource="http://comicmeta.org/vocab/Polystyrene"></cbo:encased>
  <cbo:inBox rdf:resource="#MyBoxOfComics"></cbo:inBox>
</rdf:Description>

<rdf:Description rdf:about="#MyComics" rdf:type="cbo:Collection">
  <dc:title>My Comic Book Collection</dc:title>
  <cbo:owner rdf:resource="https://sean.petiya.com/#me"/>
  <cbo:box>
    <rdf:Description rdf:about="#MyBoxOfComics">
      <dc:title>Box #1</dc:title>
    </rdf:Description>
```

```xml
    </cbo:box>
  </rdf:Description>
```

Figure 34. CBO Collection Model RDF/XML example.

The example above describes the location of a specific comic book item (Figure 34). This particular item would be inferred to be a dctype:PhysicalObject, because of the properties applied to its description, specifically those that identify a cbo:PhysicalAttribute (e.g., cbo:encased). The state, grade, and condition of an item, and the material it is contained in, all form a unique vocabulary of comic book and comic book collecting terminology.

4.1.3 CBO SKOS Vocabulary

Recognizing the need for a controlled vocabulary of comic book terminology, the individuals (owl:NamedIndividual) in the Ontology were redefined in parallel using a separate SKOS vocabulary. This solution allows for better defining the semantic relationships between terms, and potentially mapping to or incorporating existing comic book thesauri and other related projects. This process is accomplished by adding an additional rdf:type of skos:Concept to each individual, while still retaining its original class from the Ontology (Figure 35). The SKOS document is then hosted in a new namespace: http://comicmeta.org/vocab. As noted previously, a resource can be identified as a member of multiple classes, including both owl:Class and skos:Concept. This polyhierarchy illustrates a design pattern for working with OWL and SKOS proposed by Bechhofer and Miles (2008).

```xml
<NamedIndividual rdf:about="http://comicmeta.org/vocab/Polystyrene">
    <rdf:type rdf:resource="cbo:Plastic"/>
    <rdf:type rdf:resource="skos:Concept"/>
    <rdfs:label xml:lang="en">Polystyrene</rdfs:label>
</NamedIndividual>
```

Figure 35. OWL/SKOS example.

Within the SKOS file, additional relationships between terms in the vocabulary are defined using skos:narrower, skos:related, skos:broader, etc. This establishes a controlled vocabulary alongside the Ontology, and provides dereferenceable, controlled terms for instances of: *Binding*, *Comic Age*, *Condition*, *Edition*, *Format*, *Publication Frequency*, *Genre*, *Grade*, *Page Type*, *Paper*, *Plastic*, *Quality*, *Role*, and *Physical State*. An application can then load both the OWL ontology and SKOS vocabulary (e.g., by using an owl:imports statement) to utilize the additional semantic relationships defined in SKOS.

4.1.4 Content Negotiation

Communication between machines on the Web is handled by the Hypertext Transfer Protocol (HTTP) protocol. Using HTTP, when a Semantic Web application attempts to dereference a URI, it can specify the type of content it wishes to receive back using the *Accept* request header. This header tells a machine what type of a data to send back in response to a request for content. Most HTTP clients, such as a web browser, will request HTML by default. Content negotiation for both the CBO OWL ontology and SKOS vocabulary is handled by a custom web application written in C# and ASP.NET, and hosted in a Microsoft Internet Information Services (IIS) environment. All URIs in both the OWL (http://comicmeta.org/cbo/) and SKOS (http://comicmeta.org/vocab/) namespaces are dereferenceable, and respond to content requests for HTML, RDF/XML, Turtle, JSON, N3, and N-Triples. RDFa has also been embedded into the HTML response for each term. Additionally, each rendered HTML document specifies links to any alternative content types available (Figure 36).

```
<link rel="alternate" type="application/rdf+xml" href="/cbo/Comic.rdf" />
<link rel="alternate" type="text/json" href="/cbo/Comic.json" />
<link rel="alternate" type="text/turtle" href="/cbo/Comic.ttl" />
<link rel="alternate" type="text/rdf+n3" href="/cbo/Comic.n3" />
<link rel="alternate" type="text/rdf+nt" href="/cbo/Comic.nt" />
```

Figure 36. Links to alternate versions of a document embedded in HTML.

Conversion from the RDF/XML encoding of the OWL specification to the requested content type is facilitated by the dotNetRDF library (dotnetrdf.org). It is possible to demonstrate the content negotiation process by simulating a Semantic Web application using a UNIX shell and the cURL utility:

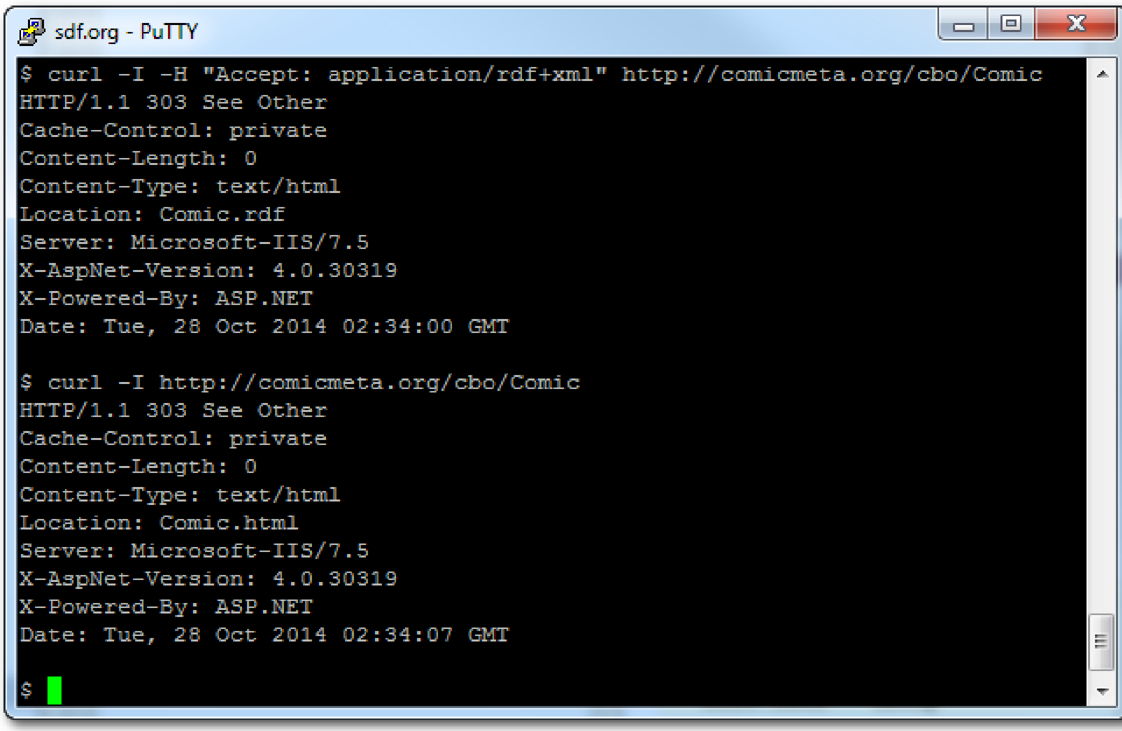

Figure 37. Debugging content negotiation using cURL.

In the first command: *curl –I –H "Accept: application/rdf+xml" http://comicmeta.org/cbo/Comic*, a request for RDF content is made by sending the proper request header to the server. The server responds with a HTTP code of 303 ("See Other") and the location of the machine-processable RDF data (Comic.rdf). In the second command: *curl –I http://comicmeta.org/cbo/Comic*, the default header is sent and the server responds with a 303 to human-readable HTML content (Comic.html). The most important component here being that the URI *http://comicmeta.org/cbo/Comic* (note: without a file extension) is intended to refer to a *concept*—not a networked resource (see 4.1.1; see also Berners-Lee et al., 2005).

4.1.5 Human-Readable Documentation

In addition to providing machine-processable semantic data, a Web vocabulary also requires human-readable documentation that can be rendered in a web browser (W3C, 2008). In order to produce this human-readable version of the Ontology, a series of XSLT stylesheets are used to transform the RDF/XML serialization of the specification into an HTML document. Semantic annotations are added to the document using RDFa, and additional formats are converted dynamically using the dotNetRDF library (dotnetrdf.org). Similar to the content negotiation process for vocabulary terms, an application may request an RDF/XML, Turtle, JSON, N3, or N-Triples representation of the Ontology specification itself.

A custom application handles this transformation process, and generates a duplicate copy by version number, date, and time with every manual publication of the Ontology. For example, publishing the 0.2.3 version of the Ontology using this process archived two files: (a) 20140504-1524.html, and (b) 20140504-1524.rdf, and then placed these files in a folder named /0.3/, along with other 0.3.* versions of the specification. The number 1524 in this example represents a combination of hours and minutes in Coordinated Universal Time (UTC). This process allows changes to the Ontology specification to be compared and understood not just across time—which is particularly helpful during a long day of development—but also between deployed versions.

4.1.6 Semantic Versioning

The Ontology attempts to follow the Semantic Versioning Specification (SemVer) established by Tom Preston-Warner (semver.org). This document very clearly describes an understandable versioning system and defines a *major*, *minor*, and *patch* incremental structure (e.g., 0.0.0), where a *major* increment may break compatibility with previous major versions; *minor* increments denote backwards compatible changes; and *patch* increments represent backwards compatible fixes. Most importantly, when using SemVer as a guide, these changes to version numbers "convey meaning about the underlying code and what has been modified from one version to the next" (Preston-Warner, n.d., para. 5). Understanding what the incremental versioning structure of a project indicates is important for users that have implemented that vocabulary in existing applications and need to be aware of any breaking changes that may result from upgrades or modifications. When combined with the approach described in the previous section, a user can request a specific, archived version of the specification when necessary, such as when having to temporarily patch a broken system resulting from an incompatible change in the main namespace (i.e., an upgrade to a new major version).

4.2 CBO Metadata Application Profiles (AP)

While the classes and properties defined in the Ontology model a conceptualization of the comic book domain, the specification does not explicitly identify the core elements or the common pattern of data shared by all participants in the description of a comic book. The Ontology also does not identify which additional properties may be used by those participants to describe their unique perspective of the comic book, or how to address any of the functional requirements of their applications. Furthermore, the Ontology does not *need* to meet all of these requirements, such as describing an ecommerce product or topic of a blog post, because many of these concepts are fully modeled in existing Web vocabularies. However, in order to implement the vocabulary as it has been described thus far, a user would have to (a) navigate the specification in its entirety to determine which models best match the nature of their data (e.g., a publication, story, product, document, etc.); (b) learn what properties are necessary to connect these entities; (c) determine how to appropriately fill in any functional gaps using the correct classes and properties from other Web vocabularies; and (d) identify which Linked Open Data (LOD) resources to use if they want to publish Linked Data. While this may seem an exciting challenge for information professionals, for webmasters, content creators, independent researchers, historians and a host of others identified by this study as creating detailed, descriptive data, this task—while not impossible—is certainly burdensome and problematic, and is definitely a distraction. In order to improve the usability of the Ontology, the specification was modularized into a series of metadata application profiles (AP), which for the purposes of this study are limited to one Core Profile, and a set of User Profiles for agents identified in the system review (publishers, retailers, collectors, libraries, researchers). Each of these APs are described in detail in the following sub-sections, and are presented alongside guidelines and recommendations for their use in publishing Linked Data using HTML and RDFa. Examples included in this chapter have been validated for their ability to generate meaningful RDF triples from HTML using commonly available RDFa validation tools.

4.2.1 Overview and Definitions

An application profile (AP) is defined as an "assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema," which are used to express extensibility, modularity, and can be developed to meet the functional requirements of a particular application (Duval et al., 2002). The usage of APs "ensures a similar basic structure with common elements, while allowing for varying degrees of depth and detail for different user communities" (Zeng & Qin, 2008, p. 112). For the purposes of discussion, an *application* in the context of this study is a Web information system built and maintained by a particular segment of a specific user community. An AP document can be used to enforce cardinality, identify value restrictions, such as the use of a controlled term list or a particular encoding, or identify other relationships and dependencies between elements. Additionally, an AP document can include guidelines for use, clarified definitions, syntax examples, or any other components that facilitate the usage and usability of that particular profile. In this regard, an AP document is very much a method of improving the usability and ultimately the user experience of implementing a metadata vocabulary. Aligned with Nielsen's model of usability attributes (1993, p. 26), a metadata AP can improve *learnability*, *efficiency* and *memorability* by presenting a selected subset of elements from a given vocabulary along with annotated examples, which may lead to reduced *errors* and improved *satisfaction* overall. In a discussion of the difficulties inherent to publishing structured data on the Web, Pohorec et al. note that "[…] common vocabularies alleviate many problems. However the content creator has to be aware of the vocabularies and how to use them [*sic*]" (2013, p. 260). The development of an AP presents a methodology for not just improving the learnability of one vocabulary, but may also provide reference for the usage of other vocabularies alongside its own documentation.

The APs defined by this study are intended to guide implementation of the Ontology as a method for publishing Linked Data in HTML using RDFa. In order to achieve this, a subset of elements is defined with each component represented by a descriptive label, property, inverse, domain, and range, with recommendations for obligation (mandatory or repeatable), encoding format, and controlled vocabulary, and finally an example of the anticipated value.

4.2.2 CBO Core AP

The Core AP represents a common set of elements that can be used to define a comic book resource at all the levels of description identified in the study's domain model (see 3.2.4.1). It has been derived from the core XML schema developed in the pilot study, and adapted for the publication of RDFa data. The elements composing the core XML schema, while capable of containing a URI in their value space, are primarily intended to carry literal data, and the majority of these elements are connected to data properties in the Ontology. In order to better facilitate the publication of RDF, object properties describing the relationships between entities have been added to the Core AP document (Table 3). These core elements include: *publisher*, *publisherName*, *imprint*, *imprintName*, *country*, *language*, *format*, *series*, *seriesTitle*, *seriesYear*, *volume*, *volumeNumber*, *issue*, *issueNumber*, *coverImage*, *publicationDate*, *edition*, *printing*, *variance*, *copy*, *condition*, *grade*, and *certNumber*.

Table 3. CBO Core Application Profile (AP)

Label	Property	Inverse	Domain	Range	Encoding	Vocab
Publisher	publisher		Series	Publisher		VIAF; LC
Publisher Name	publisherName		Publisher	Literal		
Imprint	imprint	imprintOf	Series	Imprint		VIAF; LC
Imprint Name	imprintName		Imprint	Literal		
Country	country		Series	URI	ISO-3166	TGN; LC
Language	language		Series	URI	ISO-639-2	LC
Format	format		Publication	Format		CBV; AAT
Series	series	seriesOf	Comic	Series		
Series title	seriesTitle		Series	Literal		
Series year	seriesYear		Series	Date	ISO-8601	
Volume	volume		Series	Volume		
Volume Number	volumeNumber		Volume	Literal		
Issue	issue		Volume	Issue		
Issue Number	issueNumber		Issue	Literal		
Cover Image	coverImage		Issue	URI		
Publication Date	publicationDate		Issue	Date	ISO-8601	
Edition	edition		Issue	Edition		CBV
Printing	printing		Issue	Literal		
Variance	variance		Issue	Literal		
Copy	Copy		Issue	Copy		
Condition	condition		Issue; Copy	Condition		CBV
Grade	grade		Issue; Copy	Grade		CBV
Certification Number	certNumber		Copy	Literal		

The Core AP defines the subset of elements (`property`) from the vocabulary (`vocab`), provides a label, specifies the inverse property where necessary, and identifies both the domain (`typeof`) and range (`resource`) of each property. When this range is a literal value it can be represented by text in the HTML document, and where the expected type is an entity or URI, the value can be represented by `src` (images), `href` (links), or `property` attributes where appropriate. The profile suggests recommended encodings for formatting literal values, and vocabularies for locating LOD identifiers. In the hypertext version of this document, the values link to their respective systems or documents. Recommended obligations and example values have been removed for formatting purposes. The following snippet of markup (Figure 38) presents the study's Example B (Spider-Man) described in HTML using the Core AP.

```
<div vocab="http://comicmeta.org/cbo/" typeof="Comic">
    <div property="series" typeof="Series Volume"
          resource="http://www.comics.org/series/1570">
        <h1 property="seriesTitle">The Amazing Spider-Man (<span
          property="seriesYear">1963</span>)</h1>
        <div property="publisher" typeof="Publisher"
          resource="http://viaf.org/viaf/152451185">
            <h2 property="publisherName">Marvel Comics</h2>
        </div>
        Country: <span property="country"
          resource="http://id.loc.gov/vocabulary/geographicAreas/n-us">US</span>
        Language: <span property="language"
          resource="http://id.loc.gov/vocabulary/iso639-2/eng">English</span>
        Format: <span property="format"
          resource="http://comicmeta.org/vocab/ComicBook">Comic Book</span>
        Volume Number: <span property="volumeNumber">1</span>
        <div property="issue" typeof="Issue"
          resource="http://www.comics.org/issue/44703">
          Issue Number: <span property="issueNumber">302</span>
          Publication Date: <span property="publicationDate">1988-07</span>
        </div>
    </div>
</div>
```

Figure 38. Core AP example in HTML/RDFa.

In the above snippet, the *Comic* described is part of a series represented by a resource which embodies properties of being both a series and volume, a reality described as part of the study's content analysis (see 3.5). To accomplish this in RDFa, the `typeof` attribute is assigned two classes (e.g., `typeof="Series Volume"`). Also, while the snippet above is simulating using URIs from the GCD for the *Series*, *Volume*, and *Issue*, a content creator can very well use identifiers from any system, especially their own, or even none at all. It is just as valid to use blank nodes (b-nodes), or fragmented (hashed) identifiers in HTML/RDFa. While both solutions presented will produce valid RDF data, the "correct" solution is outside the scope of this study.

```html
<div vocab="http://comicmeta.org/cbo/" typeof="Comic">
    <div property="series" typeof="Series" resource="#AmazingSpiderMan_1963">
        <h1 property="seriesTitle">The Amazing Spider-Man</h1>
        <div property="issue" typeof="Issue" resource="#AmazingSpiderMan_1963_302">
            Issue Number: <span property="issueNumber">302</span><br />
        </div>
    </div>
</div>
```

Figure 39. Core AP example in HTML/RDFa using fragmented identifiers.

```html
<div vocab="http://comicmeta.org/cbo/" typeof="Comic">
    <div property="series" typeof="Series">
        <h1 property="seriesTitle">The Amazing Spider-Man</h1>
        <div property="issue" typeof="Issue">
            Issue Number: <span property="issueNumber">302</span><br />
        </div>
    </div>
</div>
```

Figure 40. Core AP example in HTML/RDFa using blank nodes.

The URIs of the resources identified in Figure 39 would contain the namespace of the primary document, then would terminate with the specified fragmented identifier (e.g., http://example.org/#AmazingSpiderMan_1963). In Figure 40, automatically generated identifiers would be assigned to each resource when extracted from the HTML markup (e.g., _:123456789).

4.2.3 CBO User-Specific AP

The Core AP presented in the preceding section identifies the shared set of elements, entities, and properties, or the most common pattern of data, used to describe a comic book. Understandably, different segments of this user community will require additional metadata elements to meet the functional requirements of their applications and to address the dimension or perspective of the comic book which they are describing, such as that of a product, auction item, library holding, etc. In order to better address these concerns, a set of User Profiles have been developed that contain an additional subset of elements from the Ontology and other Web vocabularies intended to address the needs identified as part of this study's system review and content analysis. These metadata application profiles (AP) are designed for: (a) publishers, (b) retailers, (c) collectors, (d) libraries, and (e) researchers. They are *alpha versions* based on a preliminary analysis of systems and content, and would most likely be modified if subsequent research was conducted to solicit feedback, acquire additional analysis, etc. However, the long-term benefit to this approach is that the concerns, goals, and objectives identified for each segment of the user community can be addressed through their respective AP document or documents, without requiring breaking changes to be made to the Ontology itself. Additionally, because each User AP is dependent upon the Core Profile as a base, each remains entirely interoperable and compatible with the core model of the Ontology (Figure 41), regardless of any changes or additions made to its AP document.

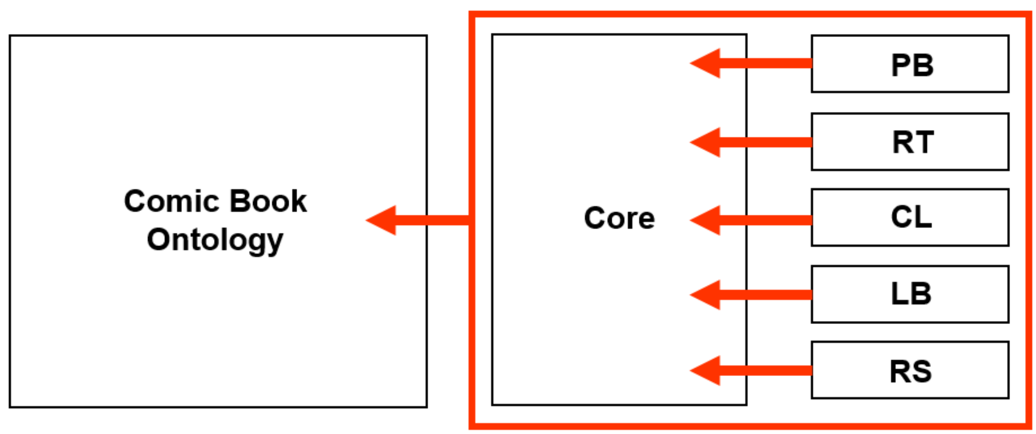

Figure 41. Model of CBO metadata application profiles (AP).

Not only does this approach potentially improve the usability of the vocabulary for each user group, it also protects the usability of each adjacent group by not requiring that group to adopt any additional elements, vocabularies, etc. that do not match their functional requirements. This process of layering application profiles meets the requirements of ensuring "a similar basic structure with common elements" (Zeng & Qin, 2009, p. 112), while achieving "[…] interoperability with the original base schemas" (Duval et al., 2002). Groups requiring more prescriptive APs—perhaps for exchanging data between closed systems—can develop localized profiles derived from these documents.

4.2.3.1 CBO Publisher (PB) AP

The Publisher (PB) AP addresses some of the following primary goals and requirements as identified in the results of the study: (a) displaying details about a comic's artist, writer, and cover artist; (b) advertising variants; (c) announcing an on-sale date; (d) identifying a rating or age guidelines; and (e) linking to a digital preview or sample. A complete list of goals and requirements identified is summarized in the PB persona (Appendix E.1). The PB-AP (Table 4) includes: *writer*, *artist*, *coverArtist*, *rating*, *variant*, *onSaleDate*, *collects*, *preview*, *isbn*, *upc*, and *distributorCode*. Additionally, this profile introduces the schema:name property and namespace (`prefix`).

Table 4. CBO Publisher (PB) AP

Label	**Property**	**Inverse**	**Domain**	**Range**	**Encoding**	**Vocab**
Writer	writer		Comic	Contributor		VIAF, ULAN
Artist	artist		Comic	Contributor		VIAF, ULAN
Cover Artist	coverArtist		Comic	Contributor		VIAF, ULAN
Name	schema:name		Contributor	Literal		
Rating	rating		Issue	URI		
Variant	variant	variantOf	Issue	Issue		
On-Sale Date	onSaleDate		Issue	Date	ISO-8601	
Collects	collects	collectedIn	Issue	Issue		
Preview	preview		Issue	Issue		
ISBN	isbn		Publication	Literal		
UPC	upc		Publication	Literal		
Distributor Code	distributorCode		Publication	Literal		

```
<div vocab="http://comicmeta.org/cbo/" prefix="schema: http://schema.org/"
       typeof="Comic">
  <div property="series" typeof="Series" resource="#AmazingSpiderMan1963">
    <h1 property="seriesTitle">
      The Amazing Spider-Man (<span property="seriesYear">1963</span>)
    </h1>
    <div property="issue" typeof="Issue" resource="#AmazingSpiderMan1963_302">
      <div>
        Issue Number: <span property="issueNumber">302</span>
      </div>
      <div property="artist" typeof="Contributor"
              resource="http://viaf.org/viaf/46886250">
        Artist: <span property="schema:name">Todd McFarlane</span>
      </div>
      <div property="writer" typeof="Contributor"
              resource="http://viaf.org/viaf/56733100">
        Writer: <span property="schema:name">David Michelinie</span>
      </div>
      <a property="preview" href="http://marvel.com/comics/issue/6708/amazing_spider-
man_1963_302">Read Sample</a>
    </div>
  </div>
 </div>
```

Figure 42. Publisher (PB) AP example in HTML/RDFa.

The above example demonstrates the usage of a selected group of properties from the PB subset (Figure 42), links to preview material, and shows how the name of an artist or writer can be displayed using a property in another Web vocabulary (schema:name). This is achieved by declaring the `prefix` for schema.org (note: the schema.org prefix is included by default in RDFa and this step is not required), and adjusting the markup from the previous example to accommodate for the additional property. RDFa Lite requires the addition of `typeof=”Contributor”`, however, an alternative approach using the full RDFa syntax would be to use the RDFa Core `about` attribute in a parent HTML element.

4.2.3.2 CBO Retailer (RT) AP

The Retailer (RT) AP is very similar to the PB subset, recall the overlapping roles of the various agents within the domain as mentioned in previous chapters. Much like the PB group, RT systems also require the ability to advertise and announce upcoming titles, storylines, products, etc. However, the distinctive requirements in the RT group include: (a) selling product, and (b) indicating a sale price. A complete list of the goals and requirements identified by the study is presented in the RT persona (Appendix E.2). In order to accomplish these goals the RT-AP (Table 5) includes the following properties: *writer*, *artist*, *coverArtist*, *variant*, *onSaleDate*, *collects*, *coverPrice*, *salePrice*, *isbn*, *upc*, and *distributorCode*. In addition to schema:name, the RT-AP also adds the schema:offers property, which has a domain of schema:Product and schema:IndividualProduct (a specific item for sale), and schema:availability which has a range of schema:ItemAvailability.

Table 5. CBO Retailer (RT) AP

Label	**Property**	**Domain**	**Range**	**Encoding**	**Vocab**
Writer	writer	Comic	Contributor		VIAF; ULAN
Artist	artist	Comic	Contributor		VIAF; ULAN
Cover Artist	coverArtist	Comic	Contributor		VIAF; ULAN
Name	schema:name	Contributor	Literal		
Variant	variant	Issue	Issue		
On-Sale Date	onSaleDate	Issue	Date	ISO-8601	
Collects	collects	Issue	Issue		
Cover Price	coverPrice	Issue	Literal	ISO-4217	
Sale Price	salePrice	Copy	Literal	ISO-4217	
ISBN	isbn	Publication	Literal		
UPC	upc	Publication	Literal		
Distributor Code	distributorCode	Publication	Literal		
Availability	schema:availability	schema:Offer	schema:		SCHEMA

			ItemAvailability		
Offer	schema:offers	schema:Product; schema:IndividualProduct	schema:Offer		

It is not the objective of the Ontology to model ecommerce, and schema.org includes a substantial product model derived from the GoodRelations (GR) ecommerce vocabulary. The concerns of the comic book domain with regard to retail overlap significantly with the wider community, and as identified in the persona document, system review, and content analysis, many RT systems include a variety of related products and merchandise, in addition to comic books. There is an economic correlation with using schema.org or GR to describe products on the Web, and the usage of these classes and properties gives the RT group the ability to markup other merchandise as well.

```
<div vocab="http://comicmeta.org/cbo/" prefix="schema: http://schema.org/"
       typeof="Comic">
    <div property="series" typeof="Series" resource="#AmazingSpiderMan1963">
        <h1 property="seriesTitle">The Amazing Spider-Man (<span
              property="seriesYear">1963</span>)
        </h1>
        <div property="issue" typeof="Issue" resource="#AmazingSpiderMan1963_302">
            <div>
                Issue Number: <span property="issueNumber">302</span>
            </div>
            <div property="copy" typeof="Copy schema:IndividualProduct">
                <div>
                    Grade: <span property="grade"
                           resource="http://comicmeta.org/vocab/9.2">9.2</span>
                    Condition: <span property="condition"
                           resource="http://comicmeta.org/vocab/NM-">NM-</span>
                    Certification: <span property="certNumber">0226722022</span>
                </div>
                <div property="schema:offers" typeof="schema:Offer">
                    Sale Price: <span property="salePrice">20.00</span>
                    Availability: <span property="schema:availability"
                                  resource="http://schema.org/InStock">In Stock</span>
                </div>
            </div>
        </div>
    </div>
</div>
```

Figure 43. Retailer (RT) AP example in HTML/RDFa.

In the example above (Figure 43), a specific comic book is described as being offered for sale using a combination of properties from the Ontology and schema.org.

4.2.3.3 CBO Collector (CL) AP

In the Collector (CL) AP, focus is placed primarily on the goals specified in the persona document, which include: (a) display the contents of a collection; (b) indicate items for sale; and (c) lookup estimated value (see Appendix E.3). Much like the overlap between the PB and RT groups, the CL group also shares many requirements with the RT group, particularly regarding the sale of items. Additionally, the CL system review and content analysis revealed that many CL systems and research databases are similar, if not hybrid systems. The CL-AP (Table 6) is the first profile to require more comic book domain specific properties, and introduces a semantically narrower set of roles: *colorist*, *inker*, *letterer*, and *penciller*. In addition to other properties, the ability to describe a *collection* and a collection *item* (including *artwork*) are added to the CL subset.

Table 6. CBO Collector (CL) AP

Label	Property	Domain	Range	Encoding	Vocab
Artist	artist	Comic	Contributor		VIAF; ULAN
Colorist	colorist	Comic	Contributor		VIAF; ULAN
Inker	inker	Comic	Contributor		VIAF; ULAN
Letterer	letterer	Comic	Contributor		VIAF; ULAN
Penciller	penciller	Comic	Contributor		VIAF; ULAN
Writer	writer	Comic	Contributor		VIAF; ULAN
Cover Artist	coverArtist	Comic	Contributor		VIAF; ULAN
Name	schema:name	Contributor	Literal		

Cover Date	coverDate	Issue	Date	ISO-8601	
Cover Price	coverPrice	Issue	Literal	ISO-4217	
Purchased Price	purchasePrice	Copy	Literal	ISO-4217	
Sale Price	salePrice	Copy	Literal	ISO-4217	
Value	value	Copy	Literal	ISO-4217	
Offer	schema:offers	schema:Product; schema:IndividualProduct	schema:Offer		SCHEMA
Page	Page	Issue	Page		
Page Number	pageNumber	Page	Literal		
Artwork	artwork	Page	Artwork		
Artwork Type	artworkType	Artwork	ArtworkType		CBV
Collection	collection	Collector	Collection		
Collection Item	item	Collection	Item		

4.2.3.4 CBO Library (LB) AP

The focus of the study's system review and content analysis regarding libraries centered on digital comic book libraries. These digital libraries represent very significant efforts to digitize public domain comic book materials, the provenance of which belongs to many comic book collectors with collections containing pre-1960s material. The task of scanning and preparing these comics is often undertaken by fans, readers, collectors, and other members of the digital library's community. The Library (LB) AP (Table 7) retains parts of previous profiles, but introduces *genre*, *owner*, *storyTitle*, and *pageCount*. The label for "Offer" has been adjusted to read "Holding" in order to eliminate any ambiguity regarding its purpose in this context. Administrative metadata, such as file size, download count, etc. have not been included in the current version of the Ontology.

Table 7. CBO Library (LB) AP

Label	**Property**	**Domain**	**Range**	**Vocab**
Writer	writer	Comic	Contributor	VIAF; ULAN
Artist	artist	Comic	Contributor	VIAF; ULAN
Name	schema:name	Contributor	Literal	

Genre	genre	Comic	URI	CBV; LC
Story Title	storyTitle	Story	Literal	
Page Count	pageCount	Story	Literal	
ISBN	isbn	Publication	Literal	
ISSN	schema:issn	schema:Periodical	Literal	
Availability	schema:availability	schema:Offer	schema: ItemAvailaility	
Holding	schema:offers	schema:Product; schema:IndividualProduct	schema:Offer	SCHEMA

While the LB-AP does not make recommendations specific to public or academic library Web information systems, the subset anticipates having applicability in those contexts, and demonstrates the use of schema:Offer to announce the availability of a holding using recommendations from the Schema Bib Extend Community Group (see 2014). Also included from schema.org are schema:issn and schema:Periodical. Important to note regarding library data on the Web is that, much like large ecommerce websites where the majority of content is non-comics material, there are more contextually appropriate Web vocabularies for libraries (e.g., BIBFRAME, RDA, etc.), which are compatible with the Ontology and include any properties necessary to meet additional functional requirements. Each AP—for all groups—is only a starting point.

```
<div vocab="http://comicmeta.org/cbo/" prefix="cbv: http://comicmeta.org/vocab/"
       typeof="Comic">
    <div property="series" typeof="Series schema:Periodical schema:PublicationVolume"
              resource="#BlackHood">
        <h1 property="seriesTitle">Black Hood</h1>
        <div property="issue" typeof="Issue schema:PublicationIssue"
              resource="http://comicbookplus.com/?dlid=22302">
            Issue Number: <span property="issueNumber">12</span>
            <div property="copy" typeof="Copy schema:Product">
                Format: <span property="format"
                          resource="cbv:DigitalComic">Digital Comic</span>
                <div property="schema:offers" typeof="schema:Offer">
                    Availability: <span property="schema:availability"
                          resource="http://schema.org/OnlineOnly">Online</span>
                </div>
            </div>
        </div>
```

```
    </div>
</div>
```

Figure 44. Library (LB) AP example in HTML/RDFa.

In the LB-AP example (Figure 44), the friendly neighborhood Spider-Man has been replaced with the licensing-friendly, public domain Black Hood, and the availability of a digital comic book at Comic Book Plus (comicbookplus.com) has been announced using schema.org following the best practices published by the Schema Bib Extend Community Group (see 2014). Additionally, this example illustrates the alignment of the Ontology's *Publication* dimension with schema:Periodical, schema:PublicationVolume, and schema:PublicationIssue.

4.2.3.5 CBO Researcher (RS) AP

The Researcher (RS) group shares the most in common with the CL group. However, whereas the principal goal of the CL group is collection management—the goals, objectives, and requirements of the RS group are more closely aligned with being able to research a specific copy, issue, story, publisher, artist, etc. When a collector needs to find the appearance of a certain character, or first work by an artist in a series, those tasks are more closely related to research. Deciding to buy or sell a copy of a comic book as result, is a by-product of that research task. Thus, the Researcher (RS) AP (Table 8) focuses on these tasks and contains a subset of properties that include: *appearance*, *character*, *team*, *story*, *story arc*, *story title*, *page*, and *sequence* for describing content; a utility property for adding *notes*; and an expanded selection of roles including the ability to define a *contributor* with any unique *role* value not specified by an existing property. The RS-AP also introduces dc:title as a mechanism for adding a title to any entity being described, and the domain for schema:name is broadened in this context so it can be applied to characters, locations, etc. Note, the range for schema:name specified by schema.org is schema:Thing, which is presumed to be semantically compatible with owl:Thing.

Table 8. CBO Researcher (RS) AP

Label	**Property**	**Inverse**	**Domain**	**Encoding**	**Vocab**
Contributor Role	role		Contributor		CBV
Name	schema:name		owl:Thing		
Artist	artist		Comic		VIAF; ULAN
Colorist	colorist		Comic		VIAF; ULAN
Editor	editor		Comic		
Inker	inker		Comic		VIAF; ULAN
Letterer	letterer		Comic		VIAF; ULAN
Penciller	penciller		Comic		VIAF; ULAN
Writer	writer		Comic		VIAF; ULAN
Contributor	contributor		Comic		
Cover Artist	coverArtist		Comic		VIAF; ULAN
Appearance	appearance		Comic		DBpedia
Character	character		Comic		DBpedia
Team	team		Comic		DBpedia
Cover Date	coverDate		Issue	ISO-8601	
On-Sale Date	onSaleDate		Issue	ISO-8601	
Cover Price	coverPrice		Issue	ISO-4217	
Page	pageNumber		Issue		
Page Number	pageNumber		Page		
Artwork	artwork	artworkOf	Page		
Artwork Type	artworkType		Artwork		CBV
Story	story		Issue		
Story Arc	storyArc		Story		
Story Title	storyTitle		Story		
Sequence	sequence		Page		
Collection	collection		Collector		
Collection Item	item	itemOf	Collection		
Note	note		owl:Thing		
Title	dc:title		owl:Thing		

While this subset does attempt to meet the requirements identified in the RS persona document (Appendix E.5), it is not entirely exhaustive and users may wish to borrow elements from one or more existing APs, or locate any additional elements necessary in the Ontology specification document.

4.3 Publishing Linked Data in HTML/RDFa

The publication of Linked Data, in which data follows a set of principles using standards such as URIs and RDF to connect to other resources, can be accomplished using two methodologies: (a) *parallel*, in a separate dataset, or (b) *inline* using semantic annotations, or structured data, in existing content. Data published in this manner is shareable, extensible, and reusable (see Baker et al., 2011), increasing its overall value and visibility. Publishing Linked Data using RDFa represents an opportunity for Web information systems, databases, and other applications to implement Semantic Web technologies using existing HTML interfaces, content, and data. The following section presents a combination of the strategies discussed in Section 4.2 to produce an example Web page which publishes structured data using the Ontology, and links that content to various Linked Open Data (LOD) resources.

4.3.1 Example Web Page Content

The following example (Figure 45) combines the strategies and solutions presented in the previous section to produce an example Web page that uses elements from the Ontology to publish Linked Data using HTML and RDFa Lite 1.1.

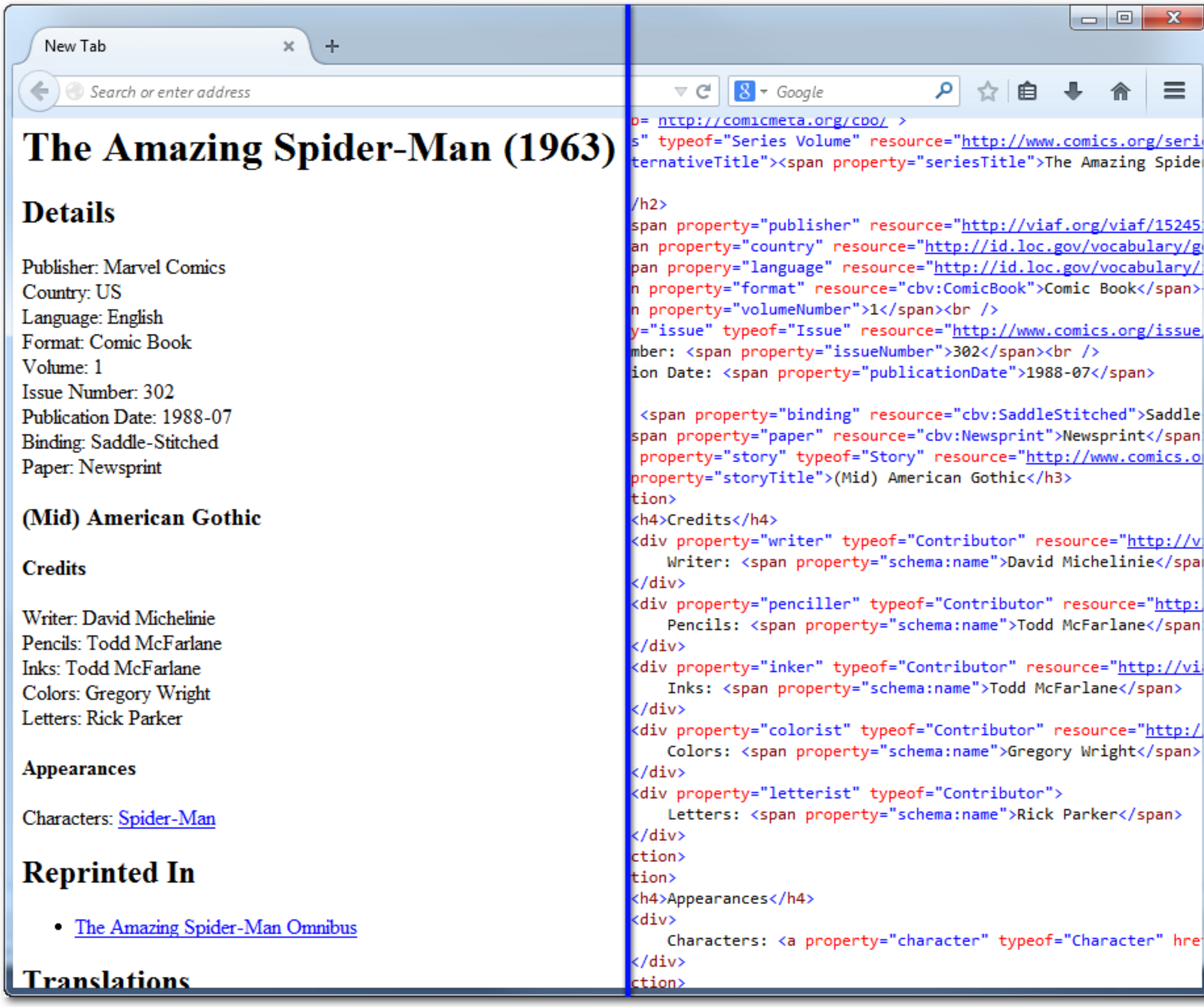

Figure 45. Linked Data in HTML/RDFa example.

This example describes all components of the study's Example B (Spider-Man), including the original issue, certified copy, translation, and the collected edition that reprints the comic. A variety of LOD URIs are used to reference the entities appearing in this record. Country and language are referenced using Library of Congress (LC) resources, the publisher and contributors are referenced using VIAF, and characters are identified using DBpedia. Comic specific terminology has been referenced using the study's SKOS vocabulary, and while the GCD URLs do not currently respond to requests for RDF data, they are still valid identifiers for comic book resources, and potentially valid LOD identifiers. The visualization below (Figure 46) was generated with RDFa Play (rdfa.info/play), using an abbreviated version of the example:

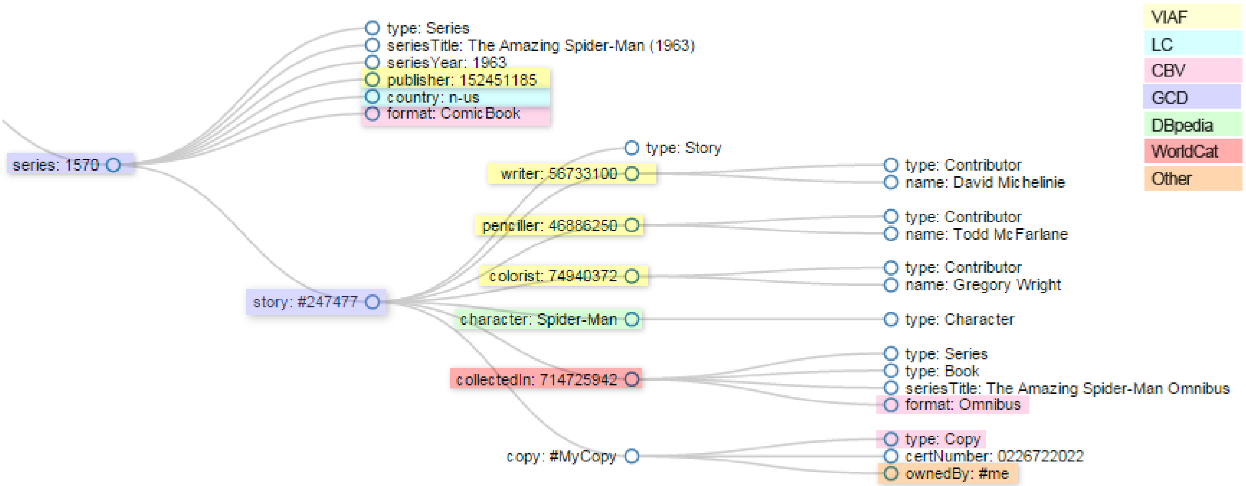

Figure 46. Linked Data in HTML/RDFa visualization.

When extracted by an RDFa parser, a total of 52 RDF triples are produced from the HTML in Figure 45. The complete HTML example is available in Appendix H and extracted RDF data is available in Appendix I. This example demonstrates the capability of using existing HTML Web content and information systems to publish Linked Data about comic books and graphic novels, as well as the potential for connecting this detailed, specialist and community created data to the larger Linked Open Data (LOD) cloud.

4.3.2 Dynamic Content Templates

Adding semantic annotations to HTML content requires familiarity with the concepts in both the vocabulary and the content being described. It is very easy to misidentify a chunk of information, incorrectly add a property, declare the wrong type, etc. While this is particularly difficult and cumbersome when marking up content by hand, when rendering dynamic content from a database, or other data source (e.g., XML record set, Web service, etc.), this process is simplified and many of the errors can be reduced once a working, validated template has been constructed (Figure 47).

```
<asp:ListView ID="ComicBook" runat="server">
    <LayoutTemplate>
        <div vocab="http://comicmeta.org/cbo/">
            <asp:PlaceHolder ID="itemPlaceholder" runat="server"></asp:PlaceHolder>
        </div>
    </LayoutTemplate>
    <ItemTemplate>
        <div typeof="Comic">
            <div property="series" typeof="Series"
                 resource='<%# string.Format("#{0}", Eval("SeriesId")) %>'>
                <h1 property="seriesTitle"><%# Eval("SeriesTitle") %> (<span
                        property="seriesYear"><%# Eval("SeriesYear") %></span>)</h1>
            </div>
        </div>
    </ItemTemplate>
</asp:ListView>
```

Figure 47. Dynamic content template example in ASP.NET.

The above example shows a sample content template written using ASP.NET. In this snippet, static code is replaced with dynamic content retrieved from a data source. The `Eval` method tells the server processing this markup to look for the specified column heading (element name, property, etc.) in the data that has been bound, or assigned, to this control. Once processed server-side, it is then sent to the client and rendered in the browser as valid HTML (Figure 48).

```
<div vocab="http://comicmeta.org/cbo/">
    <div typeof="Comic">
        <div property="series" typeof="Series" resource='#012345678'>
            <h1 property="seriesTitle">Amazing Spider-Man (<span
                    property="seriesYear">1963</span>)</h1>
        </div>
    </div>
</div>
```

Figure 48. Dynamic content rendered in HTML/RDFa.

The solution presented is just one approach to creating templates with semantic annotations, and can be adapted to web applications built using PHP, Ruby, Python, etc. Many content management systems (CMS) also allow the markup of their templates to be edited, and the approach can be extended to add structured data to websites built using Drupal, Joomla, WordPress, etc.

4.3.3 Validating and Extracting Results

All of the examples, both RDF/XML and HTML/RDFa, presented in this study have been checked for accuracy and conformance using validation services provided by the W3C, and other tools available on the Web. Additionally, all examples have been published to a Web repository for reference (sean.petiya.com/thesis). The validation and extraction of results from the HTML/RDFa examples followed a 5-step process:

1. **Validation of RDFa 1.1 Syntax**
    a. W3C RDFa Validator (http://www.w3.org/2012/pyRdfa/Validator.html)
    b. Yandex Structured Data Validator (https://webmaster.yandex.com/microtest.xml)
2. **Extraction of RDF Triples from HTML/RDFa Markup**
    a. RDFa Play (http://rdfa.info/play/)
3. **Conversion of Turtle Syntax to RDF/XML**
    a. EasyRDF Converter (http://www.easyrdf.org/converter)
4. **Validation of RDF/XML Syntax**
    a. W3C RDF Validation Service (http://www.w3.org/RDF/Validator/)
5. **Validation of Consistency with Ontology**
    a. Protégé Desktop (http://protege.stanford.edu)

This process accomplishes the following:

- **Step 1**: Validate that RDFa annotations are conformant (W3C), and that a search provider (Yandex) can surface structured data from the HTML provided.
- **Step 2, 3 and 4**: Validate an RDFa parser's (RDFa Play) ability to extract meaningful and valid (W3C) RDF triples from the HTML provided.
- **Step 5**: Validate that the RDF data extracted is consistent with the Ontology using a semantic reasoner (Protégé Desktop).

Using direct input and the validation process described above, the final HTML/RDFa example (Appendix H) was loaded into each of the tools listed. Valid RDF data was then extracted, and was confirmed to be consistent with the Ontology, meaning the results were inferred by a reasoner to be semantically accurate (i.e., a series entity is asserted to be a series, not a publisher, volcano, hotdog, etc.). It should also be noted that a simpler process for validation can be implemented by replacing the W3C RDFa 1.1 Validator with the W3C RDFa 1.1 Distiller and Parser (http://www.w3.org/2012/pyRdfa), and skipping Steps 2-3. However, this shorter method only tests a single RDFa parser, whereas the longer methodology tests multiple parsers for their ability to surface structured data from the examples tested.

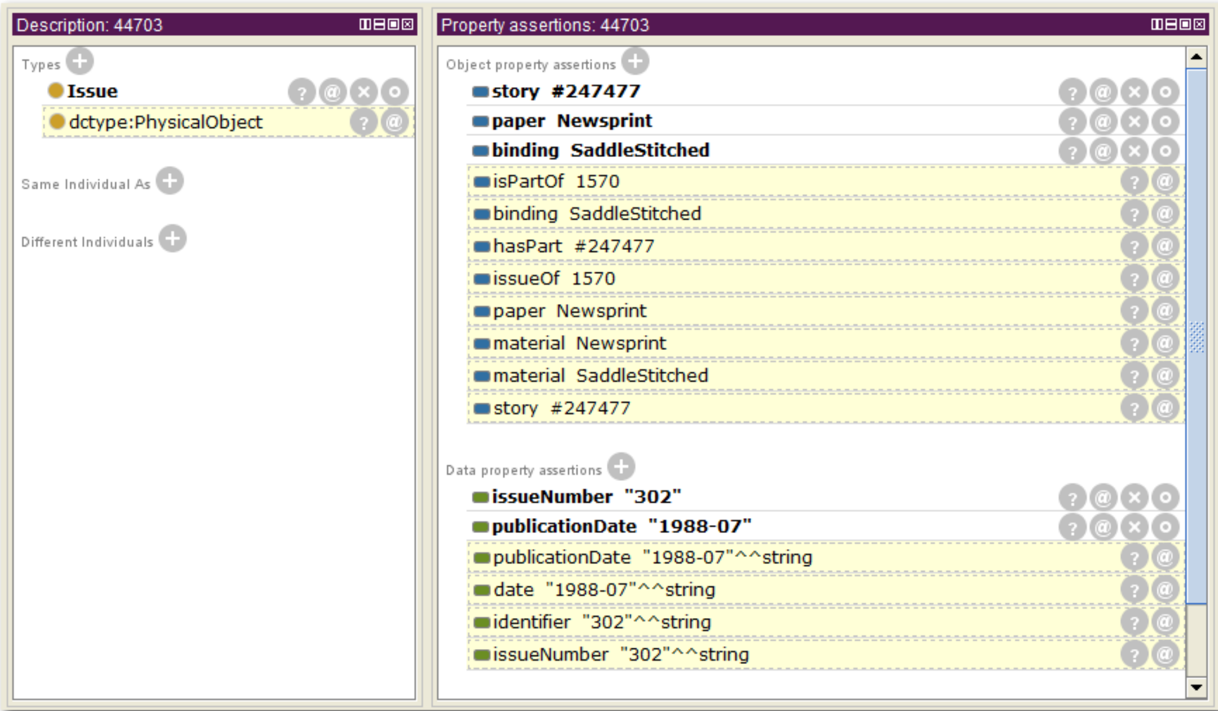

Figure 49. Consistency of HTML/RDFa example with Ontology.

The screenshot above (Figure 49), presents the inferences (assertions) made about the RDF data extracted from the final HTML/RDFa example (Appendix H) for the comic book issue being described. The implication is that it is possible to extract valid, meaningful, and semantically consistent RDF data from structured data embedded in existing Web information systems and content using HTML/RDFa that can then be interlinked to resources and data in other systems. This publication process can be guided using metadata application profile (AP) documents or other recommendations, and automated through the use of dynamic content templates. Using the common model of the CBO Ontology as a switch or map, the data extracted from existing HTML content can then be combined, repurposed, and reused across a variety of information systems and services in a Semantic Web of comics.

Chapter V

Summary and Conclusions

This study developed an ontology for comic books and comic book collections using a case study approach. A collection of information resources and example materials were used to develop both a domain and metamodel for comic books. After the initial metadata vocabulary and schema were built, a pilot study was conducted to test the structures using a workflow for converting CSV data records to XML, then to RDF/XML. The project then focused on publishing Linked Data in HTML/RDFa, and an exploration of existing information systems and content was completed. Based on the results of this analysis, the ontology was aligned with the findings and then modularized into separate metadata application profiles (AP) based on the identified needs and goals of each user group. Each AP includes a condensed subset of the ontology and selections from existing Web vocabularies designed to meet the anticipated functional requirements of each group for publishing content on the Web. Additionally, each profile inherits a core schema that represents the common elements shared amongst all participants, and the components required to successfully identify a specific comic book resource at all levels of description presented in the study's domain model (see 3.2.4.1). The following sections present the significance and limitations of the study, recommendations based on the results, and suggestions for future research.

5.1 Significance and Merits

The final result of the study is a domain ontology and metadata vocabulary for comics, titled the Comic Book Ontology (CBO). Previously, a comprehensive, extensible, open metadata specification designed specifically for describing comic books and comic book collections was not available for linking resources about comics in the Semantic Web, or in other data environments. The CBO Ontology offers a dereferenceable vocabulary for publishing Linked Data about comic books, as well as a schema for sharing, exchanging, and converting records between systems and services.

The comic book domain is vast and incorporates many participants, including publishers, distributors, retailers, libraries, archives, researchers, historians, and collectors—each having their own primary areas of concern and focus, cataloging practices, and levels of data standardization and control. Thus, information about comic books exists in various degrees of quality and completeness, and in a wide variety of catalogs, databases, and indexes, as well as other information resources and systems, especially on the Web. Comic books are serial publications with "irregular publication and frequent changes in title, character, creator and even publisher" (Fee, 2008, p. 175). And between different systems (and within the material itself), serial characteristics are not treated uniformly, with the notion of volume and series, and volume and issue often overlapping or not being represented at all (see 4.1.2.2). While the comic book in its magazine format can be most readily identified as a serial—as a multi-format medium, it does not always comfortably fit this paradigm. In order to solve this problem, the study implements a flexible, semantic data model that uses existing Web vocabularies to establish both an upper ontological model and a polyhierarchy in which concepts in the comic book domain can then be expressed as belonging to multiple superclasses (e.g., schema:Product and schema:CreativeWork, etc.), particularly where those classes have not been declared *disjoint*, meaning a resource can embody properties identified as characteristic of one or more classes without error. This allows the CBO Ontology to be placed in a larger Semantic Web environment while more accurately reflecting the reality of the material, and still retaining compatibility at a core schematic level with other data and systems within the domain. Relying on the AAA, OWA, and NUNA principles of the Semantic Web (see 2.2.1), the upper-model can be altered or replaced entirely while protecting backwards compatibility with previous implementations of the Ontology (see 3.2.4.3). Furthermore, future projects, studies, and systems can implement additional properties as necessary, define an entirely different upper-model, or even establish their own vocabulary as an extension (see 4.1.1). The Ontology, as an "explicit specification of a conceptualization" (Gruber, 1995, p. 908), is not an attempt to strictly enforce a standardization of comic book data. Rather, it is an attempt, or *one* possible conceptualization, that explores various methodologies and solutions for achieving interoperability between a variety of systems and services. Systems and applications looking for strict validation, quality control, and enhanced consistency should implement these functions at the more appropriate local or system-level (see 5.2.3 and 5.3.1 for recommendations), rather than at the conceptual level.

In order to establish a semantic data model that incorporated the various dimensions and perspectives of the comic book, a case study was developed to explore existing information resources, reference materials, and a collection of the actual materials being described (see 3.2). Additionally, publishing structured data inline using existing HTML content was determined to be a feasible approach to publishing Linked Data for many existing Web information systems. Thus, a review of these systems by user group and their content was completed using preliminary user research methodologies, notably a content analysis and a summary of that analysis in the form of persona documents which itemized the major goals and functional requirements identified. Using the personas for reference and direction, a set of metadata application profiles (AP) were designed, each of which include a subset of elements from the Ontology and other Web vocabularies designed to meet the needs and requirements of each group. This approach reduces the amount of documentation a user must uncover in order to implement a vocabulary, determine how to meet their application's functional requirements (e.g., sell a product, announce a holding, etc.), and ultimately contribute their piece of information to the knowledge graph regarding a resource. Much like Dublin Core Application Profiles (DCAP) (see Coyle & Baker, 2009), these documents can then be changed or modified to address specific needs by group without making changes to the specification, or modifying the underlying core schema. Using this Core AP as a base, additional User APs can then be developed that further describe this rich, collaborative domain, while retaining compatibility with previous profiles and existing data (see 5.4.3). Customized APs may also be stacked to express additional requirements, elements, or constraints, building layers of interoperability between systems as necessary. This research design and approach can be applied to developing Web vocabularies for similar knowledge domains and areas of specialization that are enriched or supported entirely by the efforts of independent researchers and libraries, or other dedicated user communities. Some examples include pulp magazines, music merchandise and memorabilia, fanzines, and other aspects of popular culture, even local music and art communities, which often do not become the focus of knowledge institutions until decades after their dissolution. Note, what is being described by this study is not solely a methodology for improving search engine results (see 5.3.4), rather, it is an approach for capturing and describing knowledge from multiple information systems, services, and sources—and extending its applicability.

Finally, as an RDFS/OWL ontology, the Comic Book Ontology (CBO) is compatible with a wide variety of existing resources and vocabularies. Semantic Web applications and services, especially current and future library catalogs supporting RDF, can use the Ontology to implement next-generation or social cataloging features that connect their data and collections to other systems containing information about comic book resources. The flexibility of the model as described in the previous section, would allow a library information system to retain its core vocabulary or cataloging standard (e.g.., RDA, BIBFRAME, etc.), while participating in the exchange of additional details about comic books in an open Semantic Web environment. Identifying the potential benefits and solutions presented by RDA with regard to cataloging graphic novels, Tarulli notes that next-generation catalogs “are meant to allow for easier access to collections and exploring relationships that have currently been unavailable in a library catalog” and remarks on the importance of the “development of collaborative partnerships between graphic novel experts and catalogers” in enhancing the visibility and accessibility of those collections (2010, p. 220). The trade-off for the comic book community is access to library-quality data and enhanced bibliographic control, especially in the form of name authority provided by Linked Open Data (LOD) resources like VIAF or the LC Name Authority File, among many other Linked Data (LD) services. Not to mention the ability to locate rare or significant comics in special collections and archives, or even sought after storylines and issues reprinted in the pages of materials located in the graphic novel collection of a local library. The CBO Ontology can function as a bridge between these systems, facilitating the collaborative partnership described by Tarulli, and providing the functionality needed to define the connections between systems and resources containing rich, descriptive details about comic books and graphic novels, and their publishers, creators, stories, and characters.

5.2 Limitations

5.2.1 Interpretation of Data and Limitation of User Personas and Profiles

This study relied on preliminary user research techniques to collect and analyze data, meaning the research did not involve feedback or participation from actual users, rather it developed and confirmed assumptions using methodologies that might occur at the beginning stages of a project, specifically through a review of existing systems and an analysis of content. However, this approach is not without its merit, as it provides a cost-effective solution in contrast to traditional user studies, and the repurposing of existing research and external resources has significant benefits for the beginning stages of user experience research. Goodman, Kuniavsky, and Moed explain that by “going outside your immediate development context, you can quickly acquire a high level perspective that you probably wouldn’t be able to create [internally]” and “although it may involve some amount of interpretation and extrapolation, a perusal of high-level data may quickly focus [research] by eliminating some obvious problems” (2012, p. 405). Additionally, a competitive analysis or review of existing applications and systems benefits from analyzing the end-result of any user research those services had previously invested in their products. Admittedly, however, this does require an interpretation of those results and assumptions of what the conclusions may have been which led to them. This study would benefit significantly from additional research based on user feedback and analysis, especially in determining and measuring the perceived usability of the study’s products (i.e., the Ontology and application profiles), perhaps by measuring their ability to be implemented successfully without error. Because the user persona documents and resulting application profiles (AP) were developed from this methodology, they too would benefit from additional, confirmative user research and feedback. As “skeletons”, the personas are living documents that support this kind of iterative research and can be refined or even split into narrower personas as new insights are discovered, or expanded upon to summarize additional research and major findings.

5.2.2 Usage Rights and Licensing

One very important aspect of publishing structured data not addressed by this study is the licensing of that content, or the usage rights extended to the various parties in an open exchange of data. Presumably, any system that elects to markup data in this manner, or publish it in a parallel dataset, exposes that content to a variety of services, applications, crawlers, indexers, etc., which will consume it. While how that data may ultimately be used is unpredictable, groups looking to participate in an open exchange should make licensing information clear. Publishers and consumers alike should also consider that, because Linked Data relies heavily on identifiers, their URI (i.e., the domain name) may be viewed as a form of attribution. Also, because the comic book domain is largely for-profit, stakeholders will need to weigh the benefits of publishing their data in this manner with any potential liabilities, business goals, etc. However, the benefits of the Linked Data and Linked Open Data (LOD) approach are significant and those interested should consult the Library Linked Data Incubator Group’s Final Report, which summarizes the benefits for libraries, many of which can be extended to other institutions (e.g., data that is shareable, extensible, and reusable) (see Baker et al., 2011; see also W3C, 2014a). It is important to note that publishing semantic content in HTML does not necessarily expose that content to crawlers more so than non-semantic content. HTML, like XML, is machine-parsable making it subject to *scraping*, or content harvesting. Institutions looking to benefit from structured data, while also protecting their intellectual property on the open Web, may have their interests best served by proactively developing and announcing open data licensing policies.

5.2.3 Validation and Quality of RDF Data

While this study discusses validation throughout, data is considered to be "valid" if it was syntactically valid, meaning it was conformant with standards like RDF and RDFa. The issue of validation and quality at the local, or system-level, is not addressed. However, systems looking to impose constraints on data exchanged in an open environment can build application profiles (AP), implement constraint languages like XML Schema or RDF Data Shapes, or develop similar documentation to express requirements and address quality expectations. One potential solution would be to take a defined AP (e.g., the Ontology's PB-AP; see 4.2.3.1), and adapt that for a more precise exchange of data. This would effectively create multiple layers of profile documents that would succeed in achieving interoperability between systems at various levels of constraint and quality, while still retaining compatibility with the core schema (see Figure 41). Introducing restrictions at the conceptual level reduces the flexibility and extensibility of the model in an open environment like the Semantic Web. Baker et al. identify that "defining integrity constraints separately from RDF vocabularies results in flexible vocabularies with greater reusability" and that a vocabulary "[…] is more reusable the fewer constraints it defines" (Baker et al., 2014, p. 579). If constraints were imposed by the CBO Ontology directly in its specification, the exchange of data between systems and extensibility described in Section 5.1 would not be possible. Instead, participants can apply validation rules at the system-level, perhaps dropping any systems deemed to have poor quality data from future queries.

5.2.4 Publisher Bias in Example Materials

The researcher recognizes the bias present in the materials collected during the initial phases of the research (see 3.2.2), and while the availability of the items had much to do with it, also admits to a life-long case of "Marvel Mania" as a possible cause. While generalized results and findings would most likely be similar, others are encouraged to explore these concepts and relationships in the materials of their favorite comic book publishers, formats, and stories. And especially genre, as this study focused primarily on American superhero comic books.

5.3 Recommendations

5.3.1 Validating the Quality of Comic Book Data

As pointed out in this study, and by other researchers (see Fee, 2008; see also Fee, 2013), comic book data is often inconsistent and may vary in terms of quality and completeness. There is no established standardization used universally by the domain. However, this should not deter potential participants from publishing their data using the Ontology (or possible future extensions), or others from consuming that data. For example, using the Work Model (see 4.1.2.2) a resource may be described as having qualities of both series and volume, or a volume and issue. An application consuming or storing that graph data can then determine how best to divide it (e.g., between two or three entities in a relational model, or classes in an object-oriented environment), if at all. The most significant benefit is that the resource is identified as a comic book, where it can then be linked to systems like the Grand Comics Database (GCD), ComicVine (comicvine.com), Freebase (freebase.com), or any Wiki, where granular, detailed, and descriptive information about issues, stories, characters, and events may be published.

Validation instead can be applied at the system-level, and where precise data needs exchanged between systems, a prescriptive metadata application profile (AP) or requirements document can be established (see also 5.2.3). It is also important to note that in a string of text like "Batman Arkham Asylum 25th Anniversary Deluxe Edition TP (New Edition)" (comiclist.com), where one may see a non-conformant, lawless title, another might see, at the very least, 4-5 potential RDF triples.

5.3.2 Utilizing and Establishing Linked Open Data Identifiers

One of the immediate benefits available to systems implementing the Comic Book Ontology (CBO) or any Web vocabulary, is the ability to utilize Linked Open Data (LOD) identifiers provided by Linked Data (LD) systems and services. Not only does this methodology link resources between systems, this approach also provides an accessible form of bibliographic control, particularly subject and name authority, which would have otherwise been functionally unavailable outside of library systems. Furthermore, it allows existing information systems and services to establish their URIs as valid identifiers for comic book data. While the HTML/RDFa approach presented in this study may not solve the problem of LD content negotiation explicitly, a request for RDF data may very well be handed off (i.e., HTTP 303'd) to a parallel file format or dataset (see Figure 36), or application serving data in the format requested (see 4.1.4). Additionally, future Semantic Web applications may negotiate for content by directly extracting it from HTML documents identified as having embedded RDFa markup. This allows existing information systems to maintain their familiar HTML front-ends, while enhancing the visibility of their data using both inline semantic annotations, and parallel RDF data. It also gives the library community, and other institutions, valid identifiers for comic book resources (i.e., issues, characters, storylines, creators) and access to a wealth of other descriptive details for potential inclusion in next-generation catalogs and library information systems (see 5.1).

5.3.3 Using Dynamic Content Templates and Testing Results with Analytics

While implementing semantic annotations, or structured data, in existing HTML content may seem an arduous task at first, this study recommends embedding markup in dynamic content templates, and presents an example solution using ASP.NET (see 4.3.2). The templating approach is a common way of serving dynamic content from a database, where server-side (or client-side) directives are replaced with data from a *bound* dataset. The commonness is stressed, because not only are similar approaches available in PHP, Python, and a variety of other Web programming languages—they are also scalable to content management systems (CMS) and other web services and applications built using these languages. Software packages like WordPress, Joomla, and Drupal, which allow their HTML templates to be edited, are capable of incorporating structured data into their markup. Webmasters wary of implementing this technology may consider first developing an evaluation process.

Users can evaluate and confirm the validity of the markup in their content templates using a variety of tools from Google, Yandex, the W3C, and others (see 4.3.3). For example, once the markup has been determined to be syntactically correct, webmasters may elect to apply the semantic content template to only half of their data records (e.g., only records with odd numbered identifiers), effectively setting up a multivariate, or A/B test, where results can be compared using Web analytics. Using RDFa, webmasters can even identify types from multiple vocabularies for particular sections of content (see 4.2.3.4), in order to determine which vocabulary or combination is most appropriate for meeting their requirements. However, the overall value of search engine optimization should be weighed against the primary goals and objectives of any project considering the value of structured data. This issue is discussed further in the proceeding section (see 5.3.4), however, it should be noted that the algorithms that drive search engine results are not only proprietary, but subject to constant dynamism and change affected by many factors, such as location, prior queries, previous browsing history, etc. The same multivariate test described above, run during a different month may very well deliver entirely different results. The recommendation then is to: (a) establish a good content template with semantic markup; (b) test that markup to ensure it returns valid and meaningful structured data; (c) implement that template system-wide in a way that is measurable regarding established goals and objectives, even if that goal is just to build a good, useful website; and (d) do not be dismayed if initial results do not immediately meet expectations. Search engine indexing takes time, and is often dependent on many of the hidden factors listed above.

5.3.4 Beyond Search Engine Optimization

While the near-term gains of embedding structured data in existing HTML content using RDFa and Web vocabularies may appear to be improved search engine results, rankings, and the improved visualization of data (e.g., “rich snippets”), the potential long-term benefits should also be considered. Using a shared metadata vocabulary to structure data so that it can be exchanged, repurposed, reused, and consumed by other systems, also ensures the longevity of that information and preserves the efforts and knowledge of those projects. Additionally, it makes that data and information available to a wide variety of applications, systems, services, and devices, especially those yet to be developed or imagined. Anecdotally, if we were to consider the implementation of an ontology or metadata vocabulary 15 years ago (i.e., 1999), we might also discuss how it affected search results in HotBot and AltaVista. And 5 years prior to that (i.e., 1994), whether the visibility of that data was improved in AOL keywords, CompuServe, and the Prodigy online service (or perhaps how it affected dial-up speeds). If we go further back, maybe 10 years (i.e., 1984), whether or not it improved the findability of information in a local bulletin board system (BBS), and yet another 10 years (i.e., 1974), how the ontology or vocabulary may have affected access to the campus mainframe (see computerhistory.org/timeline). The point here being not to present an overview of computing (and perhaps compare apples to oranges), but to demonstrate that in a relatively short amount of time (40 years) digital technology can change so drastically as to be almost unrecognizable. The recommendation then is to also consider long-term benefits as well as potential future applications when determining the value of implementing and investing in Linked Data and Semantic Web technology.

5.4 Future Studies

5.4.1 Linking Resources to Digital Encodings

As identified in the study's review of related projects (see 2.1), a number of comic book markup languages exist for encoding comics content as digital documents. Notably, the Comic Book Markup Language (CBML), which extends the Text Encoding Initiative (TEI) Guidelines for Electronic Text Encoding and Interchange by providing elements for marking up panels, balloons, and captions, which represent the pictorial elements necessary for defining a visual sequence of comic art. Future studies can explore connecting comics encoded in this format to their resource descriptions in the Semantic Web, especially details about creators, characters, or stories. A future search for a rare comic could potentially return a digital representation of that work in a special collection or archive that maintains a comic book collection encoded using CBML. Additionally, the markup language as an extension of TEI supports the encoding of non-comics content, such as “editorial and news features, prose fiction, fan mail, and advertisements” (Walsh, 2012a, para. 3). Future studies may also wish to explore this documentary aspect of comics, which has particular significance for comic book historians and scholars. Much of this editorial content documented the early history of the medium, and served as an important communication channel between creators and fans well before the Internet.

5.4.2 Connecting Social, Editorial, and Scholarly Content to Comics

One of the areas not explored by the study’s system review and content analysis (see 3.4), was the descriptive relationships that exist between comic books and both editorial and social content, such as blogs, reviews, and previews, as well as articles published in journals of comics scholarship. Many existing Web vocabularies, especially schema.org, provide mechanisms for describing these content types (e.g., schema:Review, schema:Blog, schema:ScholarlyArticle, etc.). Additionally, the schema.org vocabulary provides a property for describing the subject matter of content (schema:about), and future studies may explore the identification of resources in other information systems containing data with descriptive relationships to comic books and graphic novels. Social content and the dialog between fans and creators that it creates—as evidenced in the letters pages of vintage comics—has always been an important element of comic book culture.

5.4.3 Describing Other Aspects of Comics Culture

While this study attempted to conduct a thorough review of the comic book world in an effort to develop a comprehensive domain ontology, its primary focus was on bibliographic properties and relationships. Future studies may choose to explore the wide range of elements regarding comics culture and fandom not addressed in this study or not yet incorporated into the CBO Ontology, including conventions, cosplay, fan-produced content, and relationships to other media, especially videogames and movies. Additionally, future studies may narrow their focus and develop detailed descriptions for specific comic formats (i.e., graphic novels, comic strips, or manga), publishers, or universes. Also, while it was mentioned that the Ontology's Universe Model was not an attempt at fantasy world modeling (see 4.1.2.3), and is simply intended to function as a container for these important access points to comics content, this fantasy element does have a significant role. For example, Freebase establishes both "fictional_universe" and "fictional_character" types that are used with an instance of "comic_book_character" (freebase.com). Some of these properties include powers or abilities, height, weight and medical conditions, among many others. Similarly, in *The Official Handbook of the Marvel Universe*, an index of Marvel comic book characters, properties such as alias, identity, and group affiliation are listed, as well as a "Power Grid" established for many characters which assigns a numeric value to their intelligence, strength, speed, durability, energy, and fighting skills. While these may seem like frivolous details, consider that they are not without their utility and can be used alongside other data to help creators maintain a consistent, continuous narrative that often stretches across multiple decades and titles, and in the case of Marvel Comics, now spans over 75 years of content.

5.4.4 Children's Literacy and Curriculum Development

In the Comic Book Legal Defense Fund's (CBLDF) *Raising a Reader! How Comics & Graphic Novels can Help Your Kids Love to Read*, Jaffe and Holm write that comics "[are] an incredible tool for helping to create a genuine love of reading" and that "from verbal and visual literacy to critical thinking and memory, comics are a great tool to give young readers a head start" (2014, p. 1). However, the authors note that "comics are a huge medium encompassing a wide variety of stories" and that "[t]here are graphic novels for every reader, but not every graphic novel is right for all readers" (2014, p. 13). Navigating this material can be a somewhat difficult, if not an overwhelming task. One way to help identify appropriate content is to extend the social and collaborative cataloging concepts mentioned prior (see 5.1) to identifying age appropriate materials, or comics that may help to deliver core lesson concepts and objectives. While the benefits of using comic books and graphic novels in education identified by the CBLDF are numerous, one particular item of note in this regard is that the "combination of images and short bits of text encourages understanding, particularly among ESL students and reluctant readers" (Jaffe & Holm, 2014, p. 11). Future studies may explore how to best describe the subject matter and literary themes present in comics content, especially for locating materials that address complex and delicate topics like morality, gender identity, racism and prejudice, or abuse.

5.5 Conclusion

The final result of the study is a domain ontology and metadata vocabulary for comic books and comic book collections, titled the Comic Book Ontology (CBO). It is a flexible and extensible semantic model that identifies comics using two components: (a) the form (sequential art), and

(b) the container. This approach allows the conceptualization of comics provided by the model to include comic books, comic strips, web comics, graphic novels, manga, original artwork, or any other form of sequential art most readily identified as a comic or component of a comic work, regardless of its container (i.e., a page, magazine, or hardcover book) or physicality (i.e., print or digital). The comic book domain is vast and includes multiple participants, each offering a unique perspective in the description of a comic book resource: as a product, publication, hobby, art form, etc., or a unique piece of the global graph describing the various works, concepts, relationships, and entities that compose the world of comics.

While comic book data is contained in a variety of systems, services, and other resources, it is most commonly published on the Web. In order to better facilitate the ability of various participants to contribute to the graph of comic book data using existing systems and content, a set of metadata application profiles (AP) were developed to address the goals and functional requirements of each user group. Data used to inform the design of these APs was collected during a review and analysis of existing systems and content, with major findings summarized in persona documents. Each AP developed as a result of this process builds upon a core schema, which describes a specific comic book resource at all levels of description identified in the Ontology's domain model and maintains the basic level of interoperability between profiles. This allows for a "minimal ontological commitment" (Gruber, 1995, p. 910), while allowing each group to address their unique goals, functional requirements, or validation constraints in separate APs. A series of HTML/RDFa examples were then explored, tested using common validation tools, and verified for their ability to produce meaningful, conformant RDF data that is consistent with the Ontology.

As an RDFS/OWL vocabulary the Comic Book Ontology (CBO) is compatible with a variety of Semantic Web applications and other information systems, including next-generation library catalogs implementing RDA, BIBFRAME, etc., and as an extension to these formats, offers the ability to describe additional relationships between comic books and graphic novels not previously available in library catalogs (see Tarulli, 2010). The Ontology also has potential applicability in a variety of other contexts, including special collections, archives, and museums containing comic books and related materials, and can be used to interlink information regarding the various entities, concepts, and works in the world of comics from a variety of Linked Data systems, services, and other resources—effectively building a Semantic Web of comic books and comic book collections.

APPENDICES

A. Glossary of Terms

Application Profile: A subset of elements from one or more metadata schemas, most often designed for a specific community.

Comma-Separated Values: Tabular data stored in plain-text form, also known as CSV.

Content Analysis: An analysis of existing documents, objects, and data.

Dataset: A large collection of data.

Dereferenceable: Refers to a resource that can be retrieved using an Internet protocol, such as HTTP.

Information Architecture: The organization and design of labels, content, navigation, and search functionality within a system.

Linked Data Cloud: A network of linked datasets.

Linked Data: Data that is linked to other data following a set of principles for publishing data in the Semantic Web using standards, such as URI and RDF.

Linked Open Vocabulary: Common vocabularies available for use in the description of resources.

Metadata: Data that describes other data.

Metamodel: A simplified model of a model.

Ontology: A formal representation of knowledge as a conceptualization of entities and relationships in a domain.

Persona: A sketch of a user constructed from a collection and analysis of data, used to guide the development and design of a product or service by presenting major findings or directions identified using one or multiple research methodologies.

Resource: An identifiable concept, object, document, or collection of information.

Schema: A structure, often referring to a database or metadata specification.

Schemata: Equivalent form of the term "schemas".

Semantic Data: Data from which meaning can be derived, particularly by machines.

Semantic Web: A web of data, formed by interlinked resources.

Semi-structured Data: Data that is not described by a formal model, but may be self-describing.

Structured Data: Data that is contained by a formal model, such as a vocabulary.

System Review: A review of the features and content of existing systems.

Triple: A statement used to describe a resource. Composed of three parts relating a subject via a predicate to an object.

Usability: The ease of use of a product, system, or service. May refer to learnability, efficiency, memorability, and/or satisfaction.

Unstructured Data: Data that does not have a defined model, often text or images.

Vocabulary: A formal definition of terms. Often used interchangeably with the term ontology.

B. Glossary of Comic Book Terms

Annual: A comic series typically published once per year

Back issue: A comic book issue that is no longer available on the newsstand, but may be found in a specialty store.

Balloons: A word balloon that contains the dialogue of a comic sequence.

Caption: An area of text in a comic sequence that contains narration.

Colorist: An artist responsible for applying color to comic artwork.

Comic Strip: A comic composed of a short sequence of panels, typically appearing in a newspaper.

Direct Edition: An issue of a comic book that was distributed through the direct market, or a comic book specialty store. This property is often indicated by the absence of a UPC.

Encapsulated: A comic book issue that has been enclosed in plastic and sealed by a third-party service. Encapsulation is typically part of the grading and certification process.

Graded: A comic book that has been graded by a third-party service, such as the Comic Guarantee Company (CGC). The grade represents a summary of the item's physical condition.

Graphic Novel: A collection of comic book material in book, or novel, format. A graphic novel may represent original work, or may be a collection of previously published material.

Hardcover: Used to distinguish collections of hardbound comic book material from those that are softbound, such as trade paperbacks. Hardcovers may contain additional accompanying materials, or other embellishments in their production.

Illustrator: The artist of a comic artwork, may also be referred to as *penciller*.

Inker: The artist that applies inked lines and refinement to a penciled page of comic artwork. This role may also be referred to as *finisher*, or *embellisher*.

Issue: An individual issue of a comic book series.

Letterer: The artist responsible for adding the content of word balloons, captions, and other text to a page of comic artwork.

Long Box: A storage box designed for comic books, holding roughly 300 items.

Manga: Japanese comic books.

Monthly: A comic book series published once per month, often referred to as *on-going*.

Newsstand Edition: An issue of a comic book purchased at a newsstand.

Omnibus: A large, collected edition of comic book content, often containing an entire run of a series or volume, and typically including accompanying material.

One-Shot: A comic book published only once, and not part of a monthly series.

Panel: A frame in a visual sequence of comic artwork that contains pictorial content, including word balloons, captions, and other illustrations.

Page: A page of comic artwork, containing its visual sequence and narrative content.

Raw: A comic book that has not been encapsulated or professionally graded.

Reprint: A comic book that reprints a previously published comic book.

Series: A comic book series or title, most often composed of multiple serialized issues and typically divided into identifiable volumes.

Short Box: A storage box designed for comic books, holding roughly 150 items.

Story Arc: A storyline composed of multiple parts often spanning multiple issues of the same series, or other titles in the same comic book universe.

Trade Paperback: A collection of previously released comic book material.

Variant: A comic book that is different from the regular release of an issue, may refer to cover artwork or edition.

Volume: A group of related issues published within a comic book series, often indicated by the year the volume first began publication.

Web Comic: A comic published exclusively on the Web.

Writer: The writer of a comic book story, also known as the *scripter* or *plotter*.

C. Systems Review

C.1. Information Systems Reviewed

Key	Name	Description	URL
PB	**Publishers**	**Organizations responsible for the publication of comic books and related materials.**	
PB1	Marvel	Marvel Comics, publisher of Spider-Man, Avengers, etc.	www.marvel.com
PB2	DC	DC Comics, publisher of Batman, Superman, etc.	www.dccomics.com
PB3	Image	Image Comics, publisher of The Walking Dead, Spawn, etc.	www.imagecomics.com
PB4	DarkHorse	Dark Horse Comics, publisher of Star Wars, Aliens, etc.	www.darkhorse.com
RT	**Retailers**	**Organizations responsible for sale of comic books and related items.**	
RT1	TFAW	Things From Another World, online comics retailer (Los Angeles, CA)	www.tfaw.com
RT2	MidtownComics	Midtown Comics, online comics retailer (New York, NY)	www.midtowncomics.com
RT3	Comixology	Comixology, online digital comics retailer owned by Iconology, Inc. (New York, NY)	www.comixology.com
RT4	MyComicShop	MyComicShop, online auction, marketplace, and consignment, owned by Lone Star Comics, Inc. (Arlington, TX)	www.mycomicshop.com
CL	**Collectors**	**Organizations and individuals that collect comic book materials.**	
CL1	ComicBookRealm	Collection management and price guide.	www.comicbookrealm.com
CL2	StashMyComics	Collection management.	www.stashmycomics.com
CL3	ComicCollectorLive	Price guide, marketplace, and library (database). Collection management offered in desktop client.	www.comiccollectorlive.com
CL4	ComicsPriceGuide	Collection management and price guide.	www.comicspriceguide.com
LS	**Libraries**	**Organizations responsible for the maintenance, organization, and retrieval of comic book materials.**	

LS1	ComicBookLibrary	Subscription library of public domain, digital comic books. Part of World Public Library.	www.comicbooklibrary.org
LS2	ComicBookPlus	Library of public domain comic books, digitization project, and community.	www.comicbookplus.com
LS3	DigitalComicMuseum	Library of public domain, digital comic books, and community.	www.digitalcomicmuseum.com
LS4	OpenLibrary	Initiative of Internet Archive.	www.openlibrary.org
RS	**Researchers**	**Individuals interested in the history of comic books.**	
RS1	ComicBookDb	Comic book and graphic novel database.	www.comicbookdb.com
RS2	ComicVine	Comic book and graphic novel wiki.	www.comicvine.com
RS3	GrandComicsDatabase	International indexing project and comic book database.	www.comics.org
RS4	DC-Indexes	Comic book database and research project.	www.dcindexes.com

C.2. Criteria Used for Evaluation of Systems

Key	Criteria	Description
T	**Tasks**	**Tasks available to user in system.**
T1	Edit	Edit a record in the system.
T2	Buy	Buy an item listed for sale.
T3	Sell	Sell an item.
T4	Read	Read the contents of an item.
T5	Search	Execute a basic search query.
T6	Search (Faceted)	Execute an advanced search query using facets.
T7	Browse	Discover content without entering a search query.
F	**Features**	**Additional system features.**
F1	Collection	Collection catalog management.
F2	Wishlist	Wishlist of wanted items.
F3	Checklist	Checklist of needed items.
F4	Export	Export data from the system.
F5	Import	Import data to the system.
F6	Share	Share a link to a record.
D	**Data**	**Markup of data and source code.**
D1	Structured	Presence of structured data in the form of semantic annotations, or microdata.
D2	Semantic	Semantic content present in the markup, including CSS class names.
D3	Controlled	Use of controlled term lists for either data entry, search, etc.
D4	Identifiers	Presence of identifiers that could be used as URIs (non-html endings).
D5	Links	Links to item in another system, especially ecommerce.
D6	Meta	Use of descriptive, embedded meta tags.
D7	XML	Data made available as XML.
D8	RDF	Data made available as RDF.
D9	RSS	Data made available as RSS.

C.3. System Review Results (Part 1)

	T1	T2	T3	T4	T5	T6	T7	F1	F2	F3	F4	F5	F6
PB1	0	1	0	1	1	1	1	1	1	0	0	0	1
PB2	0	0	0	0	1	1	1	0	0	0	0	0	1
PB3	0	1	0	0	1	0	1	0	0	0	0	0	1
PB4	0	1	0	1	1	1	1	1	0	0	0	0	1
	0%	75%	0%	50%	100%	75%	100%	50%	25%	0%	0%	0%	100%
RT1	0	1	0	0	1	1	1	0	1	0	0	0	1
RT2	0	1	0	0	1	1	1	0	1	0	0	0	1
RT3	0	1	0	1	1	0	1	1	1	0	0	0	1
RT4	0	1	1	0	1	1	1	0	1	0	0	0	0
	0%	100%	25%	25%	100%	75%	100%	25%	100%	0%	0%	0%	75%
CL1	1	1	1	1	1	1	0	1	1	0	1	0	1
CL2	1	0	0	0	1	1	0	1	1	1	1	1	1
CL3	1	1	0	0	1	1	1	0	1	0	0	0	0
CL4	0	1	1	0	1	1	1	1	1	1	1	0	0
	75%	75%	50%	25%	100%	100%	50%	75%	100%	50%	75%	25%	50%
LB1	0	0	0	1	1	1	1	1	0	0	0	0	1
LB2	1	0	0	1	1	1	1	0	0	0	0	0	1
LB3	0	0	0	1	1	0	1	0	0	0	0	0	0
LB4	1	0	0	1	1	1	1	0	0	0	0	0	0
	50%	0%	0%	100%	100%	75%	100%	25%	0%	0%	0%	0%	50%
RS1	1	1	1	0	1	1	1	1	1	0	1	0	1
RS2	1	0	0	0	1	1	1	0	0	0	0	0	0
RS3	1	0	0	0	1	1	0	0	0	0	0	0	0
RS4	0	0	0	0	1	1	1	0	0	0	0	0	0
	75%	25%	25%	0%	100%	100%	75%	25%	25%	0%	25%	0%	25%

C.4. System Review Results (Part 2)

	D1	D2	D3	D4	D5	D6	D7	D8	D9
PB1	0	1	0	1	0	1	0	0	0
PB2	1	1	0	1	1	1	0	0	1
PB3	0	1	0	1	1	0	0	0	1
PB4	0	1	0	1	1	0	0	0	0
	25%	100%	0%	100%	75%	50%	0%	0%	50%
PB1	0	0	1	0	0	1	1	0	1
PB2	0	1	0	0	0	1	0	0	1
PB3	0	0	0	1	0	1	0	0	1
PB4	0	1	1	0	0	0	0	0	0
	0%	50%	50%	25%	0%	75%	25%	0%	75%
PB1	1	1	1	1	1	1	0	0	1
PB2	1	0	1	0	1	1	0	0	1
PB3	0	0	0	0	0	1	0	0	1
PB4	0	1	1	1	0	1	0	0	0
	50%	50%	75%	50%	50%	100%	0%	0%	75%
PB1	1	0	0	0	0	1	0	0	0
PB2	1	0	1	0	1	1	0	0	1
PB3	0	0	0	0	0	1	0	0	0
PB4	1	1	0	1	1	1	1	1	0
	75%	25%	25%	25%	50%	100%	25%	25%	25%
PB1	0	0	1	0	1	0	0	0	0
PB2	0	1	1	1	0	1	0	0	0
PB3	0	1	1	1	1	0	0	0	0
PB4	0	0	1	0	0	0	0	0	0
	0%	50%	100%	50%	50%	25%	0%	0%	0%

D. Content Analysis

D.1. Content Analysis Queries

System	Query Description
PB	**Character and content unique to publisher.**
PB1	X-Men (X-Men #141, Days of Future Past TPB)
PB2	Superman (Action Comics #821)
PB3	Savage Dragon (Savage Dragon #198, Savage Dragon: The End TP)
PB4	Star Wars (Star Wars #20)
RT	**X-Men Days of Future Past**
RT1	X-Men Days of Future Past TPB (New Printing)
RT2	X-Men Days of Future Past TPB (New Printing) / All-New Ghost Rider #1
RT3	X-Men Days of Future Past
RT4	X-Men Days of Future Past
CL	**Uncanny X-Men #141**
CL1	Uncanny X-Men #141
CL2	Uncanny X-Men #141
CL3	Uncanny X-Men #141
CL4	Uncanny X-Men #141
LB	**Black Hood Comics/Marvels**
LB1	Black Hood Comics #12
LB2	Black Hood Comics #12
LB3	Black Hood Comics #12
LB4	Marvels (Direct Edition)
RS	**Uncanny X-Men #141**
RS1	Uncanny X-Men #141
RS2	Uncanny X-Men #141
RS3	Uncanny X-Men #141
RS4	Uncanny X-Men #141

D.2. Content Item Key

Color	Definition
Green	Structural metadata
Blue	Descriptive metadata
Purple	Administrative metadata
Orange	Links
Type	
LS	Item list page.
DI	Issue details page.
DC	Collected item details page.

D.3. Content Item Index

Item	Description	URL
PB1-DC	Collected Issue Details	http://marvel.com/comics/collection/1332/x-men_days_of_future_past_trade_paperback
PB1-DI	Issue details	http://marvel.com/comics/issue/12460/uncanny_x-men_1963_141
PB1-LS	Series details	http://marvel.com/search/?q=uncanny+x-men&category=comics&offset=0
PB2-DI	Cover image	http://www.dccomics.com/comics/action-comics-1938/action-comics-821
PB2-LS	Series details	http://www.dccomics.com/browse?series=232253&content_type=comic
PB3-DC	Issue title	https://imagecomics.com/comics/releases/savage-dragon-the-end-tp
PB3-DI	Tags	https://imagecomics.com/comics/releases/savage-dragon-198
PB4-DI	Issue details	http://www.darkhorse.com/Comics/24-049/Star-Wars-20
PB4-LS	Issue details	http://www.darkhorse.com/Comics/Browse/July+2014-September+2014---0-Z/P5wfwkt8?page=2
RT1-DC	Cover image	http://www.tfaw.com/Profile/X-Men-Days-Of-Future-Past-TPB-%28New-Printing%29___394799
RT1-LS	Item details	http://www.tfaw.com/Search?_results_use_stopwords=true&quick_sstring=days+of+future+past&_results_sstype_search=
RT2-DC	Item description	http://www.midtowncomics.com/store/dp.asp?PRID=X-Men+Days+Of+Future+Past_1164154
RT2-DI	Cover image	http://www.midtowncomics.com/store/dp.asp?PRID=All-New+Ghost+Rider+%231+Co_1339146
RT2-LS	Item details	http://www.midtowncomics.com/store/search.asp?pl=16&q=days+of+future+past
RT3-DC	Breadcrumb navigation	https://www.comixology.com/X-Men-Days-of-Future-Past/digital-comic/26355
RT3-LS	Item details	https://www.comixology.com/search?search=days+of+future+past
RT4-DC	Issue title	https://www.mycomicshop.com/search?TID=91961
RT4-LS	Item details	https://www.mycomicshop.com/search?q=days+of+future+past&pubid=10221&PubRng=
CL1-D1	Series details	http://comicbookrealm.com/series/13034/0/marvel-the-x-men-vol-1
CL1-D2	Available for sale	http://comicbookrealm.com/series/13034/0/marvel-the-x-men-vol-1
CL1-D3	Contributors	http://comicbookrealm.com/series/13034/0/marvel-the-x-men-vol-1
CL1-D4	Characters	http://comicbookrealm.com/series/13034/0/marvel-the-x-men-vol-1
CL1-D5	Collected/reprinted in	http://comicbookrealm.com/series/13034/0/marvel-the-x-men-vol-1
CL1-LS	Item details	http://comicbookrealm.com/search/comics/?a=search&series=the+x-men&method=all&x=0&y=0
CL2-DI	Issue details	http://www.stashmycomics.com/searchviewcomic.asp?kcid=56894
CL2-LS	Series details	http://www.stashmycomics.com/searchresults.asp?seriesid=29794&seriestitle=x-men&issuenumber=141&itemtype=29
CL3-DI1	Breadcrumb navigation	http://www.comiccollectorlive.com/LiveData/Issue.aspx?id=c8a2adac-7738-458e-8a33-3e52cf5d507c
CL3-DI2	Market summary	http://www.comiccollectorlive.com/LiveData/Issue.aspx?id=c8a2adac-7738-458e-8a33-3e52cf5d507c
CL3-LS	Item details	http://www.comiccollectorlive.com/LiveData/Search.aspx
CL4-DI1	Issue details	http://www.comiccollectorlive.com/LiveData/Search.aspx

CL4-DI2	For sale/wanted	http://comicspriceguide.com/titles/uncanny-x-men/141/ybqvm
CL4-LS	List/item details	http://comicspriceguide.com/comic-book-search
LB1-DI	Issue details	http://comicbooklibrary.org/eBooks/WPLBN0002098936-Black-Hood-Comics---Issue-12-by-Mlj-Archie-Comics.aspx?&Words=black%20hood
LB1-LS	List/issue details	http://comicbooklibrary.org/results.aspx?SearchEverything=black%20hood&EverythingType=0
LB2-DI	Issue details	http://comicbookplus.com/?dlid=22302
LB2-LS	Breadcrumb navigation	http://comicbookplus.com/?cid=1122
LB3-DI	Breadcrumbs/ comic details	http://comicbookplus.com/?cid=1122
LB3-LS	Breadcrumbs/ comic details	http://digitalcomicmuseum.com/index.php?cid=13
LB4-DI	Navigation/details	https://openlibrary.org/books/OL8098206M/Marvels
LB4-LS	Series details	https://openlibrary.org/works/OL15528W/Marvels
RS1-DI	Issue details/credits	http://comicbookdb.com/issue.php?ID=3454
RS1-LS	Issue details	http://comicbookdb.com/title.php?ID=60
RS2-DI	Issue details	http://www.comicvine.com/the-x-men-141-days-of-future-past/4000-20988/
RS2-LS	Series details	http://www.comicvine.com/the-x-men/4050-2133/
RS3-DI	Issue details	http://www.comics.org/issue/35059/
RS3-LS	Series details	http://www.comics.org/series/1576/
RS4-DI	Issue details	http://www.dcindexes.com/features/database.php?site=&pagetype=comic&id=48102
RS4-LS	List/issue details	http://dcindexes.com/features/database.php?site=

D.4. Publisher (PB) Content Summary

Data	PB1-LS	PB1-DI	PB1-DC	PB2-LS	PB2-DI	PB3-DI	PB3-DC	PB4-LS	PB4-DI
1	Series title	Series title	Series Title	Format	Format	Series title	Series title	Cover image	Cover image
2	Series date range	Series start year	Format	Cover image	Cover image	Issue number	Creator	Preview link	Series title
3	Format / category	Issue Number	Published date	Series title	Series title	Digital formats	Price	Series title	Issue number
4	Related image	Publication date	Rating	Issue number	Issue number	Buy link	Diamond ID	Issue number	Synopsis
5	Synopsis	Added to Marvel Unlimited	Synopsis	Date of availability	Buy link	Writer	Published date	Price	Preview link
6	Last updated date	Rating	Cover image		Date of availability	Artist	Collects (range)	Buy link	Print link
7		Writer	Print issue link		Price	Cover artist	Synopsis		Writer
8		Penciller	Digital issue link		Synopsis	Price	Print link		Artist
9		Cover Artist	Imprint		Colorist	Diamond ID	Amazon link		Colorist
10		Synopsis	ISBN		Cover colorist	Date published	BN link		Cover artist
11		Cover image	Price		Inker	Synopsis	Indie Bound link		Genre
12		Sample link	UPC		Letterer	Cover image	Digital link		Publication date
13		Print issue link	Stories		Penciller	Latest issue	Comixology link		Format
14		Digital issue link	Cover Details		Written by	Tags			Page count
15		Price	Sample link		Characters				Publication frequency
16		Imprint	Creator		Series start year				Price
17		Format	Related media		On sale date				UPC
18		Inker			Issue number				Digital link

19		Colorist			Color				
20		Letterer			Trim size				
21		Editor			Page Count				
22		Appears In							

D.5. Retailer (RT) Content Summary

Point	RT1-LS	RT1-DC	RT2-LS	RT2-DI	RT2-DC	RT3-LS	RT3-DC	RT4-LS	RT4-DC
1	Cover image	Cover image	Cover image	Cover image	Collects	Cover image	Format	Stock level	Series title
2	Series title	Series title	Series title	Series title	Page count	Series title	Publisher	Series title	Format
3	Format	Format	Format	Issue number	Rating		Series title (related)	Format	Publication year
4	Printing	Printing	Printing	Variance			Series start year	Printing	Cover image
5	Publisher	Creator	Sale price	Publisher			Series end year	Display count	Issue number
6	Publication date	Format	Cover price	Onsale date			Series title	Publisher	Printing
7	Cover price	Genre	Synopsis	Writer			Cover image	Date range	Wishlist link
8	Sale price	Publisher	Buy link	Artist			Buy link		Sale price
9	Synopsis	Publication date	Condition	Variants			Sale price		Buy link
10	Buy link	Availability		Regular price			Wishlist link		Condition
11	Wishlist link	UPC		Sale price			Collects		Publisher
12		Page count		Buy link			Synopsis		Collects
13		Synopsis		Description			Rating		Synopsis
14		Collects		Item number			Digital format		Binding
15		Cover price		Diamond code			Brand logo		Page count
16		Sale price		UPC			Writer		Cover price
17		Buy link					Artist		Item
18		Wishlist link					Genre		Grade
19							Page count		Item description
20							Print release date		Paper quality
21							Digital release date		Slab quality
22							Rating		CGC serial number

D.6. Collector (CL) Content Summary

Point	CL1-LS	CL1-DI	CL2-LS	CL2-DI	CL3-LS	CL3-DI	CL4-LS	CL4-DI
1	Series title	Series title	Series title	Series title	Cover image	Comic	Cover image	Series title
2	Volume number	Volume number	Series start year	Series start year	Series title	Issue number	Series title	Issue number
3	Series date range	Volumes start year	Series end year	Series end year	Volume number	Series title	Issue number	Volume number
4	Publisher	Volume end year	Issue number	Issue number	Format	Volume number	Series date range	Publisher
5	Issue count	Publisher	Publicatio n month	Cover image	Publisher	Format	Volume number	Cover image
6		Issue number	Publicatio n year	Synopsis	Volume start year	Publisher	Publisher	Cover image (variant)
7		Value	Story title	Descripti on	Volume end year	Publicatio n date	Country	Fact
8		Cover date	Edition	Story title	Volume issue count	Cover image	Note (Appeara nces)	Comic age
9		Cover price	Value	Story page numbers	Note	Story title	Condition	Publicatio n date
10		Searched	Price paid	Event / Story arc	Issue count	Synopsis	Condition value	Cover price
11		Owned	Condition	Writer	Cover count	Issue number (alternate)		Story arc
12		Page count	Cover image	Interior penciller	Volume start date	Variance		Grade
13		ISBN		Interior inker	Volume end date	Printing		Condition
14		UPC		Cover penciller		Cover price		Raw value
15		Est. print run		Cover inker		CCA approved		Graded value
16		Variant of		Editor		Story arc		Writer
17		Preview		Interior letterer		Contribut or		Artist
18		Cover image		Interior colorist		Role		Cover artist
19		Notes		Cover letterer		Character		Inker
20		Item		Cover colorist		Group/tea m		Letterer
21		Owner / seller		Character appearanc e		Market summary		Editor
22		Grade		Team appearanc				Colorist

				e				
23		Value		Cameo appearanc e				Item condition
24		Sale price / ask price		Character 1st appearanc e				Sale price
25		Offer / want		Team 1st appearanc e				Owner
26		Collector		Death				
27		Wanted grade		Origin				
28		Willing to pay		Resurrecti on				
29		Contribut or		Collects/r eprints				
30		Role		Collected/ reprinted in				
31		Character		Collectio n				
32		Character name		Collector				
33		Appearan ce		Contributi ng member				
34		Team / group		Age/era				
35		Universe		Publicatio n date				
36		Event		Page count				
37		Reprinted /collected in		Format				
38		Events		Value				
39		Record history		Edition				
40				UPC				
41				ISBN-10				
42				ISBN-13				
43				Accessor y/feature/ item				
44				Variants				
45				Tags				

D.7. Library (LB) Content Summary

Point	LB1-LS	LB1-DI	LB2-LS	LB2-DI	LB2-LS	LB3-DI	LB4-LS	LB4-DI
1	Series title	Cover image	Format	Issue number	Publisher	Publisher	Series title	Creator
2	Issue number	Series title	Publisher	Series title	Series title	Series title	Creator	Series title
3	Author	Issue number	Series title	Publisher	Issue number	Issue number	Descripti on	Direct edition
4	Publisher	Publisher	Related image	Cover date	Volume number	Volume number	Related image	Publicatio n date
5	Volume number	Descripti on	Volume	Language	Uploader	Filesize	Subjects	Publisher
6	Language	Book ID	Volume start date	Upload date	Date uploaded	Uploader	Cover image	Language
7	Publicatio n date	Format	Volume end date	Uploader	Filesize	Downloa d count	Publicatio n date	Format
8	Subjects	File size	Issues in collection	File size	Downloa d count	View count	Publisher	Page count
9	Collectio ns	Reproduc tion date	Notes	Page count		Preview link	Series	Dimensio ns
10	Book ID	Author	Issue number	Downloa d count		Downloa d link	Format	Weight
11	Format	Volume number	Cover date	View count		Cover image	Language	OpenLibr ary ID
12		Language	Page count	Page			Edition	ISBN-10
13		Subject	Uploader	Story title			WorldCat link	ISBN-13
14		Collectio n	Date added	Story page count			Alibris link	LibraryTh ing ID
15		Subcollec tion	File size	Feature			Amazon link	Goodread s ID
16		Publicatio n date	View count	Penciller			AbeBook s link	Worldcat link
17			Downloa d count	Inker			Biblio link	Alibris link
18				Genre			Book Depositor y link	Amazon link
19				Character			Powells link	AbeBook s link
20				Event			Record editor	Biblio link
21				Cover price				Book Depositor y link
22				Editor				Powells link
23				Sequence				Record history

D.8. Researcher (RS) Content Summary

Point	RS1-LS	RS1-DI	RS2-LS	RS2-DI	RS3-LS	RS3-DI	RS4-LS	RS4-DI
1	Series title	Series title	Series title	Series title	Series title	Series title	Publisher	Series title
2	Series year	Series year	Issue count	Issue number	Publisher	Issue number	Series title	Issue number
3	Publisher	Issue number	Publisher	Story title	Series year	Edition	Issue number	Publisher
4	Volume start date	Story title	Series year	Publisher	English	Publisher	Cover date	Cover date
5	Volume end date	Publisher	Volume number	Publicatio n date	Country	Series year	On-sale date	On-sale date
6	Country	Cover image	Descripti on	Synopsis	Volume start year	Publicatio n date		Cover price
7	English	Writer	Tracking	Cover image	Volume end year	Price		Page count
8	Notes	Penciller	Series title (alternate)	Volume	Issue count	Page count		Editor
9	Issues cataloged	Inker	Collected in	In-store date	Color	Indicia frequency		Cover penciller
10	Issue number	Colorist	Cover image	Variant cover	Dimensio ns	Indicia publisher		Cover inker
11	Story title	Letterer	Issue number	Contribut or	Paper stock	Publisher brand		Story title
12	Variant	Editor	Story title	Role	Binding	Editor		Story page count
13	Story arc	Cover artist	Publicatio n date	Character	Publishin g format	Editorial role		Feature
14	Cover date	Cover date	Contribut or	Team	Publisher brands	Color		Writer
15		Cover price	Contributi on count	Location	Indicia publishers	Dimensio ns		Plotter
16		Tagline	Character	Concept	Tracking	Paper stock		Penciller
17		Format	Appearan ce count	Object	Notes	Binding		Inker
18		Page count		Story Arc	Related image	Publishin g format		Letterer
19		Version			Issue number	Issue notes		Colorist
20		Story arc				Variant		Reprinted in
21		Synopsis				Cover image		Cover image
22		Reprinted /collected in				Variant cover art		
23		Character				Sequence title		

24		Group				Feature		
25						Sequence type		
26						Page count		
27						Penciller		
28						Inker		
29						Colorist		
30						Letterer		
31						Genre		
32						Character		
33						Reprints		
34						Contributor		
35						Role		
36						Synopsis		
37						Reprint note		

D.9. Combined Content Summary by Group

PB	RT	CL	LB	RS
Amazon link	Artist	Age	AbeBooks link	Binding
Appears In	Availability	Appearance	Alibris link	Character
Artist	Binding	Artist	Amazon link	Collected in
BN link	Brand logo	Cameo appearance	Author	Color
Buy link	Buy link	CCA approved	Biblio link	Colorist
Characters	CGC serial number	Character	Book Depository link	Concept
Collects	Collects	Character appearance	Character	Contributor
Color	Condition	Character name	Collections	Country
Colorist	Copyright	Collected/reprinted in	Cover date	Cover artist
Comixology link	Cover image	Collection	Cover image	Cover date
Cover artist	Cover price	Collector	Cover price	Cover image
Cover colorist	Creator	Collects/reprints	Creator	Cover inker
Cover details	Date range	Colorist	Date added	Cover penciller
Cover image	Description	Comic	Description	Cover price
Creator	Description	Condition	Dimensions	Description
Date of availability	Diamond code	Contributor	Direct edition	Dimensions
Date published	Digital format	Country	Download link	Edition
Diamond ID	Digital release date	Cover artist	Edition	Editor
Digital formats	Format	Cover colorist	Editor	Editorial role
Digital issue link	Genre	Cover date	Event	English
Digital link	Grade	Cover image	Feature	Feature
Dimensions	Import	Cover inker	Filesize	Format
Editor	Issue number	Cover letterer	Format	Genre
Format	Item number	Cover penciller	Genre	Group
Genre	Onsale date	Cover price	Goodreads ID	Import
Import	Page count	Death	Import	Indicia frequency
Imprint	Paper quality	Description	Inker	Indicia publisher
Indie Bound link	Print release date	Edition	ISBN-10	Inker
Inker	Printing	Editor	ISBN-13	In-store date
ISBN	Publication date	Est. print run	Issue number	Issue count
Issue number	Publication year	Event	Language	Issue number
Last updated date	Publisher	Events	LibraryThing ID	Letterer
Latest issue	Rating	Fact	Notes	Location
Letterer	Regular price	Format	OpenLibrary ID	Notes
On sale date	Sale price	Grade	Page	Object
Page count	Series end year	Graded value	Page count	On-sale date

Penciller	Series start year	Group/team	Penciller	Page count
Preview link	Series title	Import	Powells link	Paper stock
Price	Slab quality	Inker	Preview link	Penciller
Print issue link	Synopsis	Interior colorist	Publication date	Plotter
Print link	UPC	Interior inker	Publisher	Price
Publication date	Variance	Interior letterer	Related image	Publication date
Publication frequency	Variants	Interior penciller	Reproduction date	Publisher
Rating	Writer	ISBN	Sequence	Publisher brands
Related image		ISBN-10	Series	Publishing format
Related media		ISBN-13	Series title	Related image
Sample link		Issue count	Source	Reprint note
Series start year		Issue number	Story page count	Reprinted/collected in
Series title		Letterer	Story title	Reprints
Stories		Notes	Subjects	Role
Synopsis		Object	Volume	Sequence title
Tags		Offer	Volume end date	Sequence type
UPC		Origin	Volume number	Series title
Writer		Owner	Volume start date	Series year
		Page count	Weight	Story arc
		Preview	Worldcat link	Story page count
		Price paid		Story title
		Printing		Synopsis
		Publication date		Tagline
		Publication month		Team
		Publication year		Tracking
		Publisher		Variant
		Raw value		Version
		Record history		Volume
		Reprinted/collected in		Volume end date
		Resurrection		Volume end year
		Role		Volume number
		Sale price		Volume start date
		Searched		Volume start year
		Series date range		Writer
		Series end year		
		Series start year		
		Series title		
		Story Arc		
		Story arc		
		Story page		

		numbers		
		Story title		
		Synopsis		
		Tags		
		Team / group		
		Team appearance		
		Universe		
		UPC		
		Value		
		Variance		
		Variant of		
		Variants		
		Volume end date		
		Volume end year		
		Volume issue count		
		Volume number		
		Volume start date		
		Volume start year		
		Writer		

D.10. Combined Content Summary by Similar Groups (PB+RT, CL+LB+RS, CL+RS)

All Groups	**PB+RT***	**CL+LB+RS***	**CL+RS***
Cover image	Artist	Character	Character
Format	Collects	Cover date	Colorist
Issue number	Cover image	Cover image	Contributor
Page count	Creator	Cover price	Country
Publication date	Format	Description	Cover artist
Series title	Genre	Edition	Cover date
	Issue number	Editor	Cover image
	Page count	Format	Cover price
	Publication date	Inker	Description
	Rating	Issue number	Edition
	Series start year	Notes	Editor
	Series title	Page count	Format
	Synopsis	Publication date	Inker
	UPC	Publisher	Issue count
	Writer	Series title	Issue number
		Story title	Letterer
		Volume end date	Notes
		Volume number	Object
		Volume start date	Page count
			Penciller
			Publication date
			Publisher
			Reprinted/collected in
			Role
			Series title
			Story Arc
			Story title
			Synopsis
			Volume end date
			Volume end year
			Volume number
			Volume start date
			Volume start year
			Writer

* Adjusted: Administrative metadata and hyperlinks removed from data points.

E. User Persona Documents

E.1. Publisher (PB) Persona

Name	Publisher
Key	PB
Source	PB1 - Marvel Comics (www.marvel.com) PB2 - DC Comics (www.dccomics.com) PB3 - Image Comics (www.imagecomics.com) PB4 - Dark Horse Comics (www.darkhorsecomics.com)
Summary	Comic book publishers are companies responsible for the creation and publication of comic materials, including comic books, graphic novels, and other products. Many comic book publishers are the owners of the rights to many popular characters, which appear in various titles and attract a variety of reader. Publishers announce new content on their website, and may make this information available through other channels, such as social media.
Frequency	Update data on a regular basis, especially to announce new material. This typically occurs at the beginning of the week and coincides with Wednesday in the U.S., or "New Comic Book Day". May also occur with the release of related media, such as a movie, television show, or other merchandise.
Content	• Publication details for recent titles, and often back issue materials. • Credits including writer and artist, including cover art. • Samples and previews of comic books.
Location	• Website • Social media • Wiki
Goals	• Announce upcoming titles and new releases. • Advertise upcoming events and storylines. • Connect products to related media, including movies, videogames, toys and collectibles. • Link products to retail outlets, including both print and digital materials.
Requirements	• Display series title, issue number, and format. • Include details about the creator, artist, and writer. • Showcase cover art, and credits. • Link together reprints and collections of material. • Link to preview. • Link products to retail outlets where materials can be purchased. • Display rating.

E.2. Retailer (RT) Persona

<table>
<tr><td>Name</td><td>Retailer</td></tr>
<tr><td>Key</td><td>RT</td></tr>
<tr><td>Source</td><td>RT1 – Things From Another World (www.tfaw.com)
RT2 – Midtown Comics (www.midtowncomics.com)
RT3 – Comixology (www.comixology.com)
RT4 – MyComicShop (www.mycomicshop)</td></tr>
<tr><td>Summary</td><td>Comic book retailers are companies that may have both a physical and online presence, or may just have an online storefront selling either print or digital materials. Retailers typically offer both new and back issue comic books for sale, with emphasis on new inventory. This inventory may also include consigned, graded, or certified materials. It is common for comic book retailers to sell related comics merchandise, especially toys and other collectibles. Many retailers may also sell original comic artwork.</td></tr>
<tr><td>Frequency</td><td>Update data on a regular basis, especially to announce new material and coinciding with Wednesday in the U.S., or “New Comic Book Day”.
Retailers may update their website, or send an announcement to mailing list subscribers to announce a sale in which materials are discounted significantly from their cover price.</td></tr>
<tr><td>Content</td><td>• Publication details for recent titles, and often back issue materials.
• Condition and grading information for back issue or consigned materials.
• Credits including writer, artist, and cover artist.
• Links to previous issues in a series.</td></tr>
<tr><td>Location</td><td>• Website
• Social media
• Newsletter</td></tr>
<tr><td>Goals</td><td>• Announce upcoming titles and new releases.
• Advertise upcoming events and storylines.
• Display sales and promotions.
• Connect products to related media, especially toys and collectibles.</td></tr>
<tr><td>Requirements</td><td>• Display series title, issue number, and format.
• Include details about the creator, artist, and writer.
• Display condition for back issues and grade for certified items
• Showcase cover art, and credits.
• Display variant items, especially variant covers.
• Provide distinguishing information about variants, including alternate issue number and credits.
• Indicate cover or suggested retail price, and sale price.</td></tr>
</table>

E.3. Collector (CL) Persona

Name	Collector
Key	CL
Source	CL1 – ComicBookRealm (comicbookrealm.com) CL2 – StashMyComics (stashmycomics.com) CL3 – ComicCollectorLive (comiccollectorlive.com) CL4 – ComicsPriceGuide (comicspriceguide.com)
Summary	Comic book collectors are readers and fans of the genre who have developed a collection of materials overtime that includes comic books in various formats, including print and digital. These collections may also contain graphic novels, manga, and collected editions. According to Overstreet, collectors may develop their collections by publisher, creator, character, or by collecting titles and runs of a series (2013b). In addition to comics, a typical personal collection may be accompanied by original comics artwork and sketches, or other related items such as toys and collectibles. Many collectors place emphasis on physical condition and may choose to have their items graded and certified by a guarantor. A collector may occasionally offer parts or all of their collection for sale.
Frequency	May update data as new items are acquired. The database supporting many collector websites and software is continually updated and maintained through user contributions.
Content	• Details about series, title, volume, and issue. • Credits for art and story. • List of character and object appearances. • Notes about the content or credits of an issue. • Details for variant items, and reprints. • Details about condition and grade, including value. • Contents of a personal collection, including items for sale. • Specific items wanted, or needed to fill a run in a collection.
Location	• Website • Database • Social media • Desktop/Mobile Application
Goals	• Display contents of a personal collection. • Advertise items in a collection that are for sale. • Alert other users to wanted items. • Discover necessary details such as credits, appearances, or notes about a comic issue. • Lookup the estimated value of a comic issue.
Requirements	• Display series title, issue number, and format. • Include details about the creator, artist, and writer. • Identify reprints, editions, and other variants of an issue. • Provide estimated values for issues in a particular condition or grade. • Organize and identify items in a collection.

E.4. Library (LB) Persona

Name	Librarian
Key	LB
Source	LB1 – Comic Book Library (comicbooklibrary.org) LB2 – Comic Book Plus (comicbookplus.com) LB3 – The Digital Comic Museum (digitalcomicmuseum.com) LB4 – Open Library (openlibrary.org)
Summary	Comic book libraries maintain a collection of comic books and are responsible for the organization and retrieval of those materials. This collection may also include or be entirely composed or graphic novels and collected editions. Digital comic book libraries contain digitized version of public domain comic books and may provide access to these materials directly to the public, or may require users to register for service. These digitizations are often linked to known print collections of Golden Age materials, with efforts to preserve the items being conducted by members of the community supporting the digital library.
Frequency	May update data as new items are acquired. The database supporting many collector websites and software is continually updated and maintained through user contributions.
Content	• Series, title, date, edition, and creator details. • Information about holdings for print comic books and graphic novels. • Links to related websites, including LibraryThing, OCLC, etc. • Digital copies of public domain comic books. • Detailed information about series and character history. • Historical and cultural context for materials.
Location	• Website • Database • Index
Goals	• Advertise availability of items in a collection. • Announce availability of new materials. • Provide provenance information for materials. • Acquire access metadata and statistics for materials. • Align comic material and history with other contexts.
Requirements	• Organize materials by publisher and title. • Provide access to digital copies. • Cite source material for information. • Describe provenance information.

E.5. Researcher (RS) Persona

Name	Researcher
Key	RS
Source	RS1 – ComicBookDB (comicbookdb.com) RS2 – Comic Vine (comicvine.com) RS3 – Grand Comics Database (comics.org) RS4 – ComicsPriceGuide (comicspriceguide.com)
Summary	Comic book researchers are readers, collectors, and fans of the genre who have an interest in the history of the medium. Researchers may seek information about a publisher, a series, or a specific issue, in addition to many other attributes of the comic book. The focus of the research may be placed on the contents, the story, the artwork, or even the document itself, with many research databases often including information about pages of commentary, editorial, or letters appearing in a comic document.
Frequency	Ongoing and continual maintenance provided by user contributions.
Content	• Details about series, title, volume, and issue. • Credits and notes for art, stories, and sequences. • List of character, team, and object appearances. • List of variants, reprints, and translations. • Indexer notes. • Details about individual pages, or sequences in a document.
Location	• Website • Database • Index
Goals	• Provide accurate and complete data about a comic. • Include an index of characters, appearances, and events in a comic. • Encourage user participation in the maintenance of data.
Requirements	• Display series title, issue number, and format. • Showcase cover image and details. • Provide publication date and year. • Include details about the creator, artist, and writer. • Identify reprints, editions, and other variants of an issue. • List character appearances and events. • List issues by series.

F. Ontology Alignment

Superclass	Master*	Property	Type	Range
Comic	Age	comicAge	ObjectProperty	URI
ComicUniverse	Appearance	appearance	ObjectProperty	URI
Contributor	Artist	artist	ObjectProperty	URI
Agent	Author	writer	ObjectProperty	URI
Publication	Binding	binding	ObjectProperty	URI
Agent	Brand logo	schema:logo		
ComicUniverse	Cameo	cameo	ObjectProperty	URI
Publication	CCA approved	cca	DatatypeProperty	Bool
Item	CGC serial number	certNumber	DatatypeProperty	Literal
ComicUniverse	Character	character	ObjectProperty	URI
ComicUniverse	Character name	schema:name		
Publication	Collected in	collectedIn	ObjectProperty	URI
Collection	Collection	collection	ObjectProperty	URI
Agent	Collector	owner	ObjectProperty	URI
Publication	Collects	collects	ObjectProperty	URI
Comic	Color	format	ObjectProperty	URI
Contributor	Colorist	colorist	ObjectProperty	URI
Comic	Comic		Class	
ComicUniverse	Concept		Class	
Item	Condition	condition	ObjectProperty	URI
Contributor	Contributor	contributor	ObjectProperty	URI
Publication	Copyright	dc:rights		
Publication	Country	country	ObjectProperty	URI
Contributor	Cover artist	coverArtist	ObjectProperty	URI
Issue	Cover date	coverDate	DatatypeProperty	Literal
Document	Cover image	coverArt	ObjectProperty	URI
Publication	Cover price	coverPrice	DatatypeProperty	Literal
Creator	Creator	creator	ObjectProperty	URI
ComicUniverse	Death	event	Class	
Thing	Description	dc:description		
Publication	Diamond code	diamondCode	DatatypeProperty	Literal
Publication	Digital format	dct:format		
Publication	Dimensions	dimensions	ObjectProperty	URI
Publication	Direct edition	edition	Concept	
Publication	Edition	edition	ObjectProperty	URI
Contributor	Editor	editor	ObjectProperty	URI
ComicUniverse	Event	event	ObjectProperty	URI

Thing	Fact	dc:description		
Comic	Feature	feature	DatatypeProperty	Literal
Publication	Filesize	schema:fileSize		
Publication	Format	format	ObjectProperty	URI
Comic	Genre	genre	ObjectProperty	URI
Publication	Goodreads ID			
Item	Grade	grade	ObjectProperty	URI
ComicUniverse	Group	group	ObjectProperty	URI
Agent	Imprint	imprint	ObjectProperty	URI
Publication	Indicia frequency	indiciaFrequency	ObjectProperty	URI
Publication	Indicia publisher	indiciaPublisher	ObjectProperty	URI
Contributor	Inker	inker	ObjectProperty	URI
Publication	ISBN	schema:isbn		
Publication	ISBN-10	*		
Publication	ISBN-13	*		
Publication	Issue count	issueCount		
Publication	Issue number	issueNumber	DatatypeProperty	Liteal
Item	Item number	itemNumber	DatatypeProperty	Literal
Publication	Language	language	ObjectProperty	URI
Contributor	Letterer	letterer	ObjectProperty	URI
Publication	LibraryThing ID			
Item	Location			
Thing	Notes	note		
ComicUniverse	Object	object	ObjectProperty	URI
Item	Offer			
Publication	On-sale date	onSaleDate	DatatypeProperty	Date
Publication	OpenLibrary ID			
ComicUniverse	Origin	event	ObjectProperty	URI
Agent	Owner	owner	ObjectProperty	URI
Document	Page	page	ObjectProperty	URI
Document	Page count	pageCount	DatatypeProperty	
Document	Paper quality	quality	ObjectProperty	URI
Document	Paper stock	paper	ObjectProperty	URI
Contributor	Penciller	penciller	ObjectProperty	URI
Contributor	Plotter	contributor	ObjectProperty	URI
Publication	Preview	preview		
Publication	Price	price	DatatypeProperty	Literal
Item	Price paid	purchasePrice	DatatypeProperty	Literal
Publication	Printing	printing	DatatypeProperty	Literal
Publication	Publication date	publicationDate	DatatypeProperty	Date
Publication	Publication frequency	publicationFrequency	ObjectProperty	URI

Publication	Publication month			
Publication	Publication year			
Agent	Publisher	publisher	ObjectProperty	URI
Publication	Publisher brands	schema:brand	DatatypeProperty	Literal
Publication	Publishing format	publicationFrequency	ObjectProperty	URI
Publication	Rating	rating	ObjectProperty	URI
Comic	Related image			
Comic	Related media			
Publication	Reprinted in	reprintedIn	ObjectProperty	URI
Publication	Reprinted/collected in	collectedIn	ObjectProperty	URI
Publication	Reprints	reprints	ObjectProperty	URI
Publication	Reproduction date			
ComicUniverse	Resurrection	event	ObjectProperty	URI
Contributor	Role	role	ObjectProperty	URI
Item	Sale price	salePrice	DatatypeProperty	Literal
Sequence	Sequence	sequence	ObjectProperty	URI
Sequence	Sequence title	dc:title		
Sequence	Sequence type			
Publication	Series	series	ObjectProperty	URI
Publication	Series date range			
Publication	Series end year	endYear		
Publication	Series start year	startYear		
Publication	Series title	seriesTitle	DatatypeProperty	Literal
Publication	Series year	seriesYear		
Item	Slab quality			
Source	Source	dc:source		
Story	Stories	story	ObjectProperty	URI
Story	Story arc	storyArc	ObjectProperty	URI
Story	Story page count	pageCount	DatatypeProperty	
Story	Story page numbers			
Story	Story title	storyTitle	DatatypeProperty	Literal
Story	Subjects	subject	ObjectProperty	URI
Story	Synopsis	synopsis	DatatypeProperty	Literal
Story	Tagline			
Comic	Tags			
ComicUniverse	Team	team	ObjectProperty	URI
Publication	Tracking	tracking	DatatypeProperty	Literal
ComicUniverse	Universe	universe	ObjectProperty	URI
Publication	UPC			
Item	Value	value	DatatypeProperty	Literal
Publication	Variance	variance	DatatypeProperty	Literal

Publication	Variant	variant	ObjectProperty	URI
Publication	Variant of	variantOf	ObjectProperty	URI
Publication	Version	dc:hasVersion		
Publication	Volume	volume	ObjectProperty	URI
Publication	Volume end date	endDate	DatatypeProperty	Literal
Publication	Volume end year	endDate	*	*
Publication	Volume issue count	issueCount		
Publication	Volume number	volumeNumber	DatatypeProperty	Literal
Publication	Volume start date	startDate	DatatypeProperty	Literal
Publication	Volume start year	startDate	*	*
Contributor	Writer	writer	ObjectProperty	URI

*Yellow denotes a property that did not exist in the CBO Ontology prior to alignment.

G. Ontology Properties and Classes (v0.12)

Classes	Object Properties	Datatype Properties
Agent	adaptation	alternativeTitle
Artwork	adaptationOf	cca
Balloon	appearance	certNumber
Binding	artist	copyrightDate
Box	artwork	count
Building	artworkOf	coverDate
Caption	attribute	coverPrice
Character	bagged	diamondCode
Collection	balloon	distributorCode
Collector	binding	endDate
Comic	boarded	feature
ComicAge	box	imprintName
ComicUniverse	cameo	indiciaFrequency
Condition	caption	indiciaPublisher
Container	character	isbn
Contributor	collectedIn	issueCount
Copy	collection	issueNumber
Costume	collects	itemNumber
Creator	colorist	note
Distributor	comicAge	onSaleDate
Document	condition	pageCount
Edition	contains	price
Event	contributor	printing
Format	copy	publicationDate
Frequency	copyOf	publisherName
Genre	copyState	purchasePrice
Grade	country	salePrice
Guarantor	coverArt	seriesTitle
Imprint	coverArtist	seriesYear
Issue	creator	startDate
Item	distributedBy	storyTitle
Library	edition	synopsis
Material	editor	tracking
Museum	encased	upc
Object	event	value
Page	format	variance
PageType	genre	volumeNumber

Panel	grade	
Paper	guaranteedBy	
Pedigree	hasStory	
PhysicalAttribute	hasUniverse	
Place	imprint	
Planet	imprintOf	
Plastic	inBox	
Publication	inker	
Publisher	issue	
Quality	item	
Role	itemOf	
Sequence	language	
Series	letterer	
State	material	
Story	object	
Team	owner	
Thing	ownerOf	
Vehicle	page	
Volume	pageType	
	panel	
	paper	
	penciller	
	plotter	
	preview	
	publicationFrequency	
	publisher	
	quality	
	rating	
	related	
	reprintedIn	
	reprints	
	role	
	sequence	
	series	
	seriesOf	
	state	
	story	
	storyArc	
	subject	
	team	
	translation	

	translationOf	
	universe	
	universeOf	
	variant	
	variantOf	
	volume	
	volumeOf	
	writer	

*Current version available at: http://comicmeta.org/cbo

H. HTML/RDFa Content Example Full-Text

```
<!DOCTYPE html>
<html prefix="cbo: http://comicmeta.org/cbo/ cbv: http://comicmeta.org/vocab/ rdfa:
http://www.w3.org/ns/rdfa#">
<body>
    <div typeof="Comic" vocab="http://comicmeta.org/cbo/">
        <div property="series" typeof="Series Volume"
resource="http://www.comics.org/series/1570">
            <h1 property="alternativeTitle"><span property="seriesTitle">The Amazing
Spider-Man</span> (<span property="seriesYear">1963</span>)</h1>
            <section>
                <h2>Details</h2>
                Publisher: <span property="publisher"
resource="http://viaf.org/viaf/152451185">Marvel Comics</span><br />
                Country: <span property="country"
resource="http://id.loc.gov/vocabulary/geographicAreas/n-us">US</span><br />
                Language: <span property="language"
resource="http://id.loc.gov/vocabulary/iso639-2/eng">English</span><br />
                Format: <span property="format" resource="cbv:ComicBook">Comic
Book</span><br />
                Volume: <span property="volumeNumber">1</span><br />
                <div property="issue" typeof="Issue"
resource="http://www.comics.org/issue/44703">
                    Issue Number: <span property="issueNumber">302</span><br />
                    Publication Date: <span property="publicationDate">1988-07</span>
                    <br />
                    Binding: <span property="binding"
resource="cbv:SaddleStitched">Saddle-Stitched</span><br />
                    Paper: <span property="paper"
resource="cbv:Newsprint">Newsprint</span><br />
                    <section property="story" typeof="Story"
resource="http://www.comics.org/issue/44703/#247477">
                        <h3 property="storyTitle">(Mid) American Gothic</h3>
                        <section>
                            <h4>Credits</h4>
                            <div property="writer" typeof="Contributor"
resource="http://viaf.org/viaf/56733100">
                                Writer: <span property="schema:name">David
Michelinie</span>
                            </div>
                            <div property="penciller" typeof="Contributor"
resource="http://viaf.org/viaf/46886250">
                                Pencils: <span property="schema:name">Todd
McFarlane</span>
                            </div>
                            <div property="inker" typeof="Contributor"
resource="http://viaf.org/viaf/46886250">
                                Inks: <span property="schema:name">Todd McFarlane</span>
                            </div>
                            <div property="colorist" typeof="Contributor"
resource="http://viaf.org/viaf/74940372">
                                Colors: <span property="schema:name">Gregory
Wright</span>
                            </div>
                            <div property="letterist" typeof="Contributor">
```

```
                                Letters: <span property="schema:name">Rick Parker</span>
                            </div>
                        </section>
                        <section>
                            <h4>Appearances</h4>
                            <div>
                                Characters: <a property="character" typeof="Character"
href="http://dbpedia.org/page/Spider-Man">Spider-Man</a>
                            </div>
                        </section>
                        <section>
                            <h2>Reprinted In</h2>
                            <ul>
                                <li><a property="collectedIn" typeof="Comic schema:Book"
href="http://www.worldcat.org/oclc/714725942">
                                    <span property="seriesTitle">The Amazing Spider-Man
<span property="format" resource="cbv:Omnibus">Omnibus</span></span></a>
                                </li>
                            </ul>
                        </section>
                        <section>
                            <h2>Translations</h2>
                            <ul>
                                <li>
                                    <a property="translation" typeof="Comic"
href="http://www.comics.org/issue/899800">
                                        <span><span property="seriesTitle">L'Étonnant
Spider-Man (<span property="seriesYear">1993</span>)</span> #<span
property="issueNumber">5</span></span>
                                    </a>
                                </li>
                            </ul>
                        </section>
                        <section>
                            <h2>Copies</h2>
                            <ul>
                                <li>
                                    <span property="copy" typeof="Copy">
                                        <span property="guarantor"
resource="cbv:CGC">CGC</span> <span property="grade"
resource="cbv:9.4">9.4</span> #<a property="certNumber">0226722022</a> owned by
                                    <a property="ownedBy"
href="https://sean.petiya.com/#me">Sean Petiya</a> (20.00 USD)</span></li>
                            </ul>
                        </section>
                    </section>
                </div>
            </section>
        </div>
    </div>
</body>
</html>
```

I. RDF Triples Extracted from HTML/RDFa Content

```
@prefix cbo: <http://comicmeta.org/cbo/> .
@prefix cbv: <http://comicmeta.org/vocab/> .
@prefix rdfa: <http://www.w3.org/ns/rdfa#> .
@prefix schema: <http://schema.org/> .

<> rdfa:usesVocabulary cbo: .

<http://dbpedia.org/page/Spider-Man> a cbo:Character .

<http://viaf.org/viaf/56733100> a cbo:Contributor;
    schema:name "David Michelinie" .

<http://viaf.org/viaf/74940372> a cbo:Contributor;
    schema:name "Gregory Wright" .

<http://www.comics.org/issue/44703> a cbo:Issue;
    cbo:binding cbv:SaddleStitched;
    cbo:issueNumber "302";
    cbo:paper cbv:Newsprint;
    cbo:publicationDate "1988-07";
    cbo:story <http://www.comics.org/issue/44703/#247477> .

<http://www.comics.org/issue/44703/#247477> a cbo:Story;
    cbo:character <http://dbpedia.org/page/Spider-Man>;
    cbo:collectedIn <http://www.worldcat.org/oclc/714725942>;
    cbo:colorist <http://viaf.org/viaf/74940372>;
    cbo:copy [ a cbo:Copy;
            cbo:certNumber "0226722022";
            cbo:grade <http://comicmeta.org/vocab/9.4>;
            cbo:guarantor cbv:CGC;
            cbo:ownedBy <https://sean.petiya.com/#me> ];
    cbo:inker <http://viaf.org/viaf/46886250>;
    cbo:letterist [ a cbo:Contributor;
            schema:name "Rick Parker" ];
    cbo:penciller <http://viaf.org/viaf/46886250>;
    cbo:storyTitle "(Mid) American Gothic";
    cbo:translation <http://www.comics.org/issue/899800>;
    cbo:writer <http://viaf.org/viaf/56733100> .

<http://www.comics.org/issue/899800> a cbo:Comic;
    cbo:issueNumber "5";
    cbo:seriesTitle "L'Ătonnant Spider-Man (1993)";
    cbo:seriesYear "1993" .

<http://www.comics.org/series/1570> a cbo:Series,
        cbo:Volume;
    cbo:alternativeTitle "The Amazing Spider-Man (1963)";
    cbo:country <http://id.loc.gov/vocabulary/geographicAreas/n-us>;
    cbo:format cbv:ComicBook;
    cbo:issue <http://www.comics.org/issue/44703>;
    cbo:publisher <http://viaf.org/viaf/152451185>;
    cbo:seriesTitle "The Amazing Spider-Man";
    cbo:seriesYear "1963";
    cbo:volumeNumber "1" .
```

```
<http://www.worldcat.org/oclc/714725942> a cbo:Comic,
        schema:Book;
    cbo:format cbv:Omnibus;
    cbo:seriesTitle "The Amazing Spider-Man Omnibus" .

<http://viaf.org/viaf/46886250> a cbo:Contributor;
    schema:name "Todd McFarlane" .

[] a cbo:Comic;
    cbo:series <http://www.comics.org/series/1570> .
```

References

Allemang, D., & Hendler, J. A. (2008). *Semantic web for the working ontologist: Modeling in RDF, RDFS, and OWL.* Burlington, MA: Morgan Kauffman.

Antoniou, G., Groth, P., Harmelen, F.V., & Hoekstra, R. (2012). *A Semantic Web primer* (3rd ed.). Cambridge, MA: MIT Press.

Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., … Zeng, M. (2011, October 25). *Library Linked Data incubator group final report.* Retrieved from http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025

Baker, T., Coyle, K., & Petiya, S. (2014). Multi-entity models of resource description in the Semantic Web: A comparison of FRBR, RDA, and BIBFRAME. *Library Hi Tech, 32*(4), 562-582. doi:10.1108/LHT-08-2014-0081

Bechhofer, S., & Miles, A. (2008, May). *Using OWL and SKOS*. Retrieved from http://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html

Berners-Lee, T. (2009, July 18). *Linked Data.* Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Fielding, R., Masinter, L. (2005, January). *Uniform resource identifiers (URI): General syntax.* Retrieved from https://tools.ietf.org/html/rfc3986

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The story so far. *International Journal on Semantic Web and Information Systems*, *4*(2), 1-22.

Brickley, D., & Miller, L. (2014, January). *FOAF vocabulary specification 0.99*. Retrieved from http://xmlns.com/foaf/spec/

Brown, D. (2011). *Communication design: Developing website documentation for design and* *planning* (2nd ed.). Berkeley, CA: New Riders.

Cole, W. M. (2013). Preservation and storage of comic books. In J. C. Vaughn & M. Huesman (Eds.), *The Overstreet guide to collecting comics* (pp. 64-66). Hunt Valley, MD: Gemstone Publishing.

Coyle, K. & Baker, T. (2009, May 18). *Guidelines for Dublin Core application profiles*. Retrieved from http://dublincore.org/documents/profile-guidelines/

DCMI. (2004, September 10). *Library application profile.* Retrieved from http://dublincore.org/documents/library-application-profile/

DenVog, LLC. (n.d.). *CoMet specification*. Retrieved from http://www.denvog.com/comet/comet-specification

Duncan, R., & Smith, M. J. (2013). *The power of comics: History, form, & culture.* New York: Bloomsbury Academic.

Duval, E., Hodgins, W. Sutton, S. & Weibel, S. (2002). Metadata principles and practicalities. *D-Lib Magazine*, *8*(4). doi:10.1045/april2002-weibel

Eisner, W. (2008). *Comics and sequential art: Principles and practices from the legendary cartoonist.* New York, NY: W.W. Norton.

Fagan, B. D., & Fagan, J. C. (2011). *Comic book collections for libraries*. Santa Barbara, CA: Libraries Unlimited.

Fee, W. T. (2008). Do you have any Ditko? Comic books, MARC, FRBR and findability. *Serials Review*, *34*(3), 175-189. doi:10.1016/j.serrev.2008.06.003

Fee, W. T. (2013). Where is the Justice League? Graphic novel cataloging and classification. *Serials Review*, *39*(1), 37-46. doi:10.1016/j.serrev.2013.02.004

Gilliland, A. J. (2008). Setting the stage. In M. Baca (Ed.). *Introduction to metadata: Online Edition* (3rd ed). Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html

Goodman, E., Kuniavsky, M., & Moed, A. (2012). *Observing the user experience: A practitioner's guide to user research*. Waltham, MA: Morgan Kaufmann.

Google. (2013, September 26). Providing structured data. In *Custom search*. Retrieved from https://developers.google.com/custom-search/docs/structured_data

Grand Comics Database. (2012a, October 12). *Data distribution*. Retrieved from http://docs.comics.org/wiki/Data_Distribution

Grand Comics Database. (2012b, October 5). What is the GCD? In *General FAQ*. Retrieved from http://docs.comics.org/wiki/General_FAQ

Griffin, J. K. (1998). A brief glossary of comic book terminology. *Serials Review*, *24*(1), 71-76. doi:10.1016/S0098-7913(99)80104-X

Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies, 43*(5-6), 907-928.

Gruber, T. (2008). Collective knowledge systems: Where the social web meets the Semantic Web. *Web Semantics: Science, Services, and Agents on the World Wide Web, 1*(6), 4-13. doi:10.1016/j.websem.2007.11.011

He, W., Mihara, T., Nagamori, M., & Sugimoto, S. (2013). Identification of works of manga using LOD resources: An experimental FRBRization of bibliographic data of comic books. In S. Downie, R. H. McDonald, T. W. Cole, R. Sanderson, & F. Shipman (Eds.), *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 253-256). doi:10.1145/2467696.2467731

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). Functional requirements for bibliographic records. *IFLA Series on Bibliographic Control*, *19*. Retrieved from http://www.ifla.org/publications/functional-requirements for-bibliographic-records

Jaffe, M., & Holm, J. L. (2014). *Raising a reader! How comics & graphic novels can help your kids love to read!* New York, NY: Comic Book Legal Defense Fund.

Lefèvre, P. (2010). European comics. In *Encyclopedia of Comic Books and Graphic Novels*. (Vol. 1, pp. 185-192). Santa Barbara, CA: Greenwood Press.

Library of Congress. (2013, April). *Comic book collection*. Retrieved from http://www.loc.gov/rr/news/coll/049.html

McCloud, S. (1994). *Understanding comics: The invisible art.* New York, NY: HarperPerennial.

McCloud, S. (2006). *Making comics*. New York, NY: HarperCollins.

McIntosh, J. (2001, April 18). *ComicsML: A simple markup language for comics.* Retrieved from http://www.xml.com/pub/a/2001/04/18/comicsml.html

McIntosh, J. (2011, September 11). *ComicsML – XML for digital comics*. Retrieved from http://comicsml.jmac.org/

Morozumi, A., Nomura, S., Nagamori, M., & Sugimoto, S. (2009). Metadata framework for manga: A multi-paradigm metadata description framework for digital comics. In *Proceedings from International Conference on Dublin Core and Metadata Applications* (pp. 61-70). Seoul, KR: Dublin Core Metadata Initiative.

Morville, P., & Rosenfeld, L. (2007). *Information architecture for the World Wide Web* (3rd ed.).

Sebastopol, CA: O'Reilly.

MusicBrainz. (n.d.). *About*. Retrieved from https://musicbrainz.org/

Nielsen, J. (1993). *Usability engineering.* Boston: Academic Press.

NISO. (2004). *Understanding metadata.* Bethesda, MD: NISO Press. Retrieved from http://www.niso.org/publications/press/UnderstandingMetadata.pdf

O'Leary, S. (2013). Comics retailers upbeat on... *Publishers Weekly*, *260*(30), 32-35.

O'Nale, R. (2010). Manga. In *Encyclopedia of Comic Books and Graphic Novels*. (Vol. 2, pp. 378-387). Santa Barbara, CA: Greenwood Press.

Overstreet, R. M. (2013). *The Overstreet comic book price guide* (43rd ed.). Hunt Valley, MD: Gemstone Publishing.

Overstreet, R. M., & Vaughn, J. C. (2013a). *The Overstreet guide to collecting comic & animation art.* Hunt Valley, MD: Gemstone Publishing.

Overstreet, R. M., & Vaughn, J. C. (2013b). *The Overstreet guide to collecting comics.* Hunt Valley, MD: Gemstone Publishing.

Pastierovic, R. (n.d.). *Advanced comic book format.* Retrieved from https://launchpad.net/acbf

Petiya, S. (2014). Converting personal comic book collection records to Linked Data. In W. Moen & A. Rushing (Eds.), *Proceedings of the International Conference on Dublin Core and Metadata Applications: Metadata Intersections: Bridging the Archipelago of Cultural Memory* (pp. 222-224). Austin, TX: Dublin Core Metadata Initiative. Retrieved from http://dcevents.dublincore.org/IntConf/dc-2014/paper/view/290/278

Pohorec, S., Zorman, M., & Kokol, P. (2013, November). Analysis of approaches to structured data on the web. *Computer Standards & Interfaces*, *36*(1), 256-262. doi:10.1016/j.csi.2013.06.003

Preston-Warner, T. (n.d.). *Semantic versioning 2.0.0*. Retrieved from http://semver.org/

Pruitt, J., & Adlin, T. (2010). Persona conception and gestation. In C. Wilson (Ed.), *User experience re-mastered: Your guide to getting the right design* (pp. 155-219). Burlington, MA: Morgan Kaufmann.

Pyles, C. (2013, February 25). *It's no joke: Comic and collection development.* Retrieved from http://publiclibrariesonline.org/2013/02/its-no-joke-comics-and collection-development

Rhode, M., & Bullough, J. (2009). *Comics research bibliography*. Retrieved from http://homepages.rpi.edu/~bulloj/comxbib.html

Ronallo, J. (2012, February 3). HTML5 microdata and schema.org. *Code4Lib Journal*, *1*(16). Retrieved from http://journal.code4lib.org/articles/6400

Schema Bib Extend Community Group. (2014, June 13). *Holdings via offer*. Retrieved from https://www.w3.org/community/schemabibex/wiki/Holdings_via_Offer

Schema.org. (n.d.). *Frequently asked questions*. Retrieved from https://schema.org/docs/faq.html

Schema.org. (2014, May 14). *Data model*. Retrieved from https://schema.org/docs/datamodel.html

Scott, R. W. (1998). A practicing comic-book librarian surveys his collection and his craft. *Serials Review*, *24*(1), 49-56. doi:10.1016/S0098-7913(99)80102-6

Segaran, T., Taylor, J., & Evans, C. (2009). *Programming the Semantic Web.* Sebastopol, CA: O'Reilly.

Serchay, D. S. (1998). Comic book collectors: The serials librarians of the home. *Serials Review*, *24*(1), 57-70. doi:10.1016/S0098-7913(99)80103-8

Tarulli, L. (2010). Cataloging and problems with Dewey: Creativity, collaboration, and compromise. In Weiner, R. G. (Ed.), *Graphic novels and comics in libraries and archives: Essays on readers, research, history, and cataloging* (pp. 213-221). Jefferson, NC: McFarland & Company.

Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative Inquiry, 17*(6), 511-521.

Walsh, J. A. (2012a, May 17). *Comic book markup language*. Retrieved from http://dcl.slis.indiana.edu/cbml

Walsh, J. A. (2012b). Comic book markup language: An introduction and rationale. *Digital Humanities Quarterly*, *6*(1). Retrieved from http://www.digitalhumanities.org/dhq/vol/6/1/000117/000117.html

Wikipedia. (2013, October 24). *Semantic Web stack.* Retrieved from http://en.wikipedia.org/wiki/Semantic_Web_Stack

W3C. (2008, August 28). *Best practice recipes for publishing RDF vocabularies*. Retrieved from http://www.w3.org/TR/swbp-vocab-pub

W3C. (2009, November 12). *Semantic Web frequently asked questions.* Retrieved from http://www.w3.org/2001/sw/SW-FAQ

W3C. (2012a, December 12). *OWL 2 Web ontology language document overview (second edition).* Retrieved from http://www.w3.org/TR/owl2-overview

W3C. (2012b, June 7). *RDFa Lite 1.1*. Retrieved from http://www.w3.org/TR/rdfa-lite/

W3C. (2013a). *Vocabularies*. Retrieved from http://www.w3.org/standards/semanticweb/ontology

W3C. (2013b, October 29). *HTML5 Microdata*. Retrieved from http://www.w3.org/TR/microdata

W3C. (2013c, August 22). *RDFa 1.1 primer – second edition: Rich structured data markup for Web documents*. Retrieved from http://www.w3.org/TR/rdfa-primer

W3C. (2013d, August 22). *RDFa Core 1.1: Syntax and processing rules for embedding RDF through attributes* (2nd ed.). Retrieved from http://www.w3.org/TR/rdfa-syntax/

W3C. (2014a, January 14). *Best practices for publishing Linked Data*. Retrieved from http://www.w3.org/TR/ld-bp

W3C. (2014b, February 25). *RDF Schema 1.1.* Retrieved from http://www.w3.org/TR/rdf-schema

W3C. (2014c, June 2014). *RDF 1.1 primer.* Retrieved from http://www.w3.org/TR/rdf11-primer

W3C. (2014d, June 23). *WebSchemas/PeriodicalsComics*. Retrieved from http://www.w3.org/wiki/WebSchemas/PeriodicalsComics

Zeng, M. L., & Qin, J. (2008). *Metadata*. New York, NY: Neal-Schuman Publishers.