

Converting Personal Comic Book Collection Records to Linked Data

Sean Petiya | School of Library and Information Science | Kent State University, USA

ABSTRACT

The **Comic Book Ontology** (Fig. 1) is a metadata vocabulary designed for the description of **comic books** and **comic book collections**. This poster explores applying a subset of the vocabulary to personal collection records. The intention of this core application profile is to enhance the usability of the ontology by improving its understandability and identifying the minimal amount of commitment necessary to identify a specific resource. This data is commonly contained in spreadsheets, and a workflow is described for mapping data columns to XML, then applying an XSLT stylesheet to generate RDF/XML, replacing common terms with Linked Data URIs.

The comic book domain is vast. Collectors represent one of the most passionate segments of the user population and have been largely responsible for preserving the history of the medium. Many collectors have made significant contributions to comics research, scholarship, and librarianship. Linked Data represents an opportunity for this user group to share their knowledge and highlight the efforts of many research projects and collections.

BACKGROUND

Comic books are magazine format publications containing an illustrated, serialized story that occurs across multiple titles, volumes, and issues. These objects function simultaneously as a **publication**, **document**, **artwork**, and **story**. These stories compose an intricate universe of characters and events. In the library, the comic book (Fig. 2) (A) most often takes the form of a **graphic novel** or **trade paperback**, which are often collections of reprinted material. Other forms, such as the Japanese **manga**, may vary in format and presentation, while the European comic book is very similar to the North American version (B). However, despite varying forms and publication formats, all find their way into personal libraries and collections. These collections often contain items that are preserved, or “bagged and boarded”, using archival quality materials. Collectors owning rare and valuable items may choose to have their materials certified by a guarantor (C).

CORE PROFILE

The core application profile represents only the subset of elements in the **Comic Book Ontology** required to uniquely identify a comic book, and the most common pattern of data shared between user groups. The ontology's model acknowledges that a comic can be described from multiple perspectives and at various levels, including the structure of the document and its sequence of images, captions, and word balloons. It is designed to be flexible, allowing a data publisher to contribute information at any level, or create their own using new or existing vocabularies. The core profile isolates the most common and recognizable pattern, the publication model, and reduces the vocabulary to 15 elements (Table 1).

Core Profile Elements		
Element	Mandatory	Repeatable
publisherName	Yes	Yes
imprintName	No	Yes
country	Yes	No
seriesTitle	Yes	No
seriesYear	No	No
volumeNumber	No	No
issueNumber	Yes	No
language	Yes	No
format	Yes	Yes
edition	No	Yes
printing	No	No
variance	No	Yes
condition	No	Yes
grade	No	No
certNumber	No	No

Table 1 - Core Profile Elements

This core profile also utilizes a comic book domain model (Fig. 3) that aligns with a work, expression, manifestation, and item (WEMI) view of the resource. Comic books are often translated, reprinted, or collected in other works. Additionally, each comic book issue can represent a different edition, printing, or variance, all of which determine the specific manifestation being described. Both condition and grade may also be used to refer to a manifestation, rather than a specific item. This core set of elements is made available in the **Comic Book Schema (Core)**.



Fig. 2 - Comic books (Original, translated, and certified)

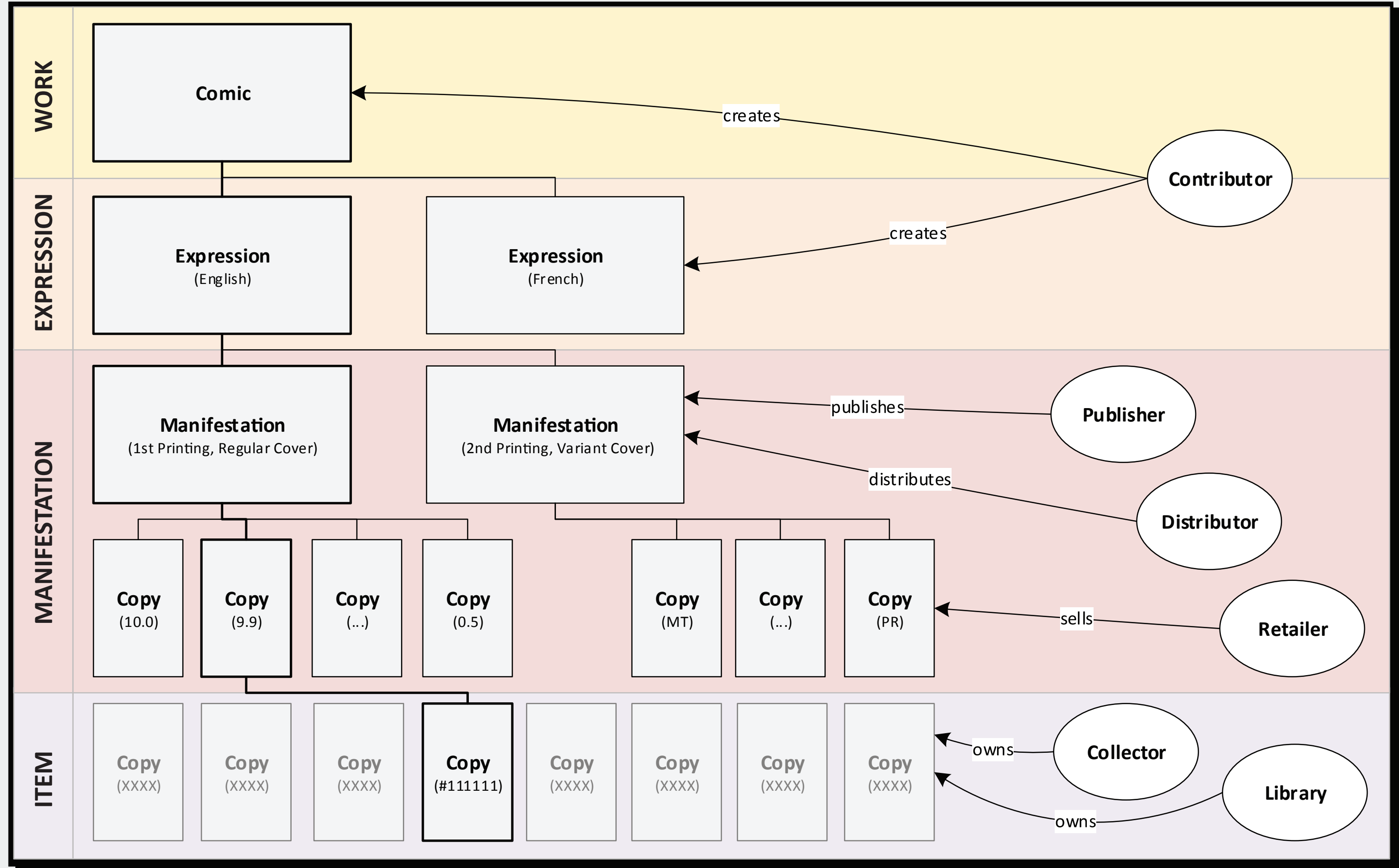


Fig. 3 - Comic book domain model

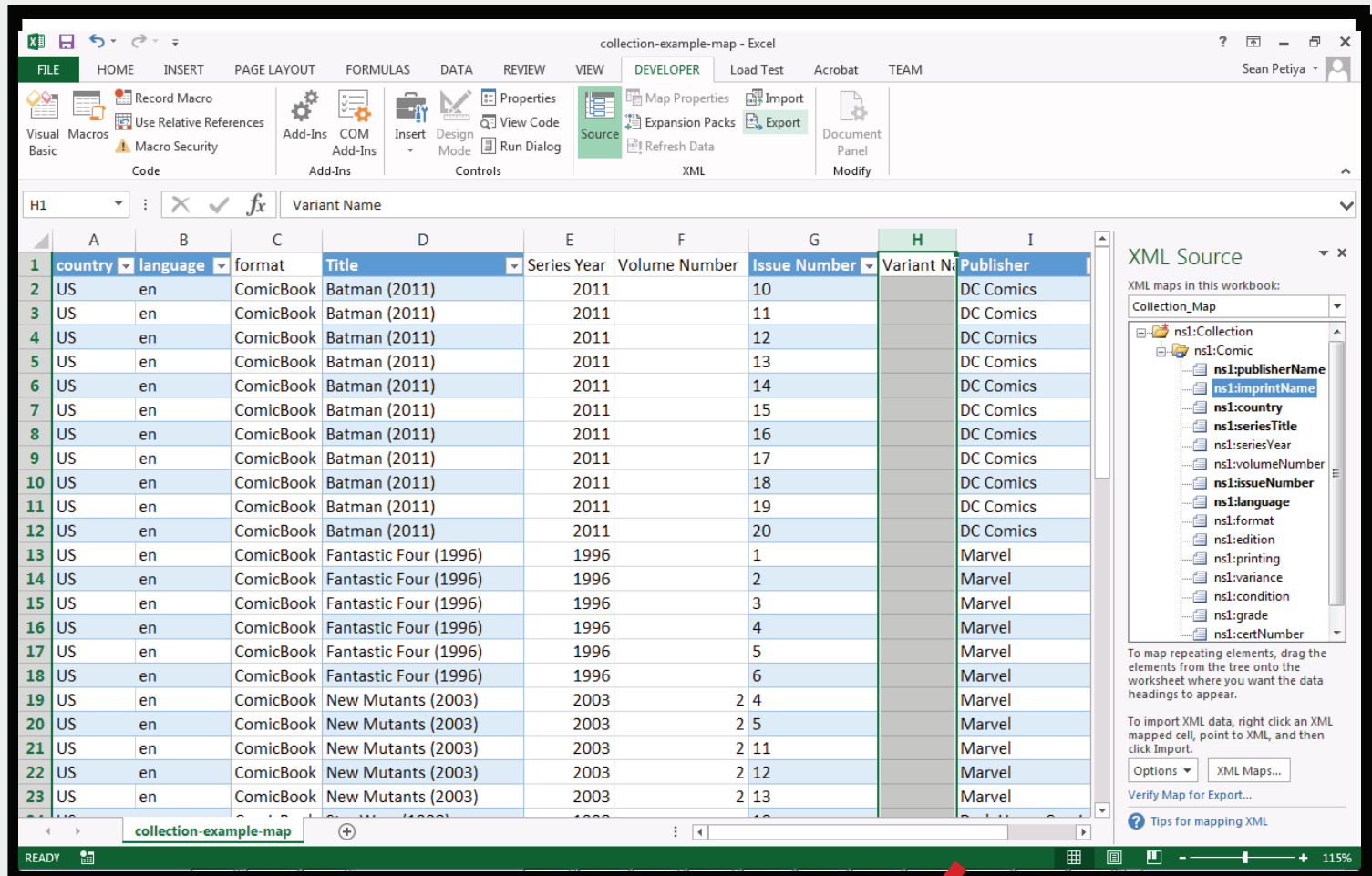
WORKFLOW

1. MAP TO XML

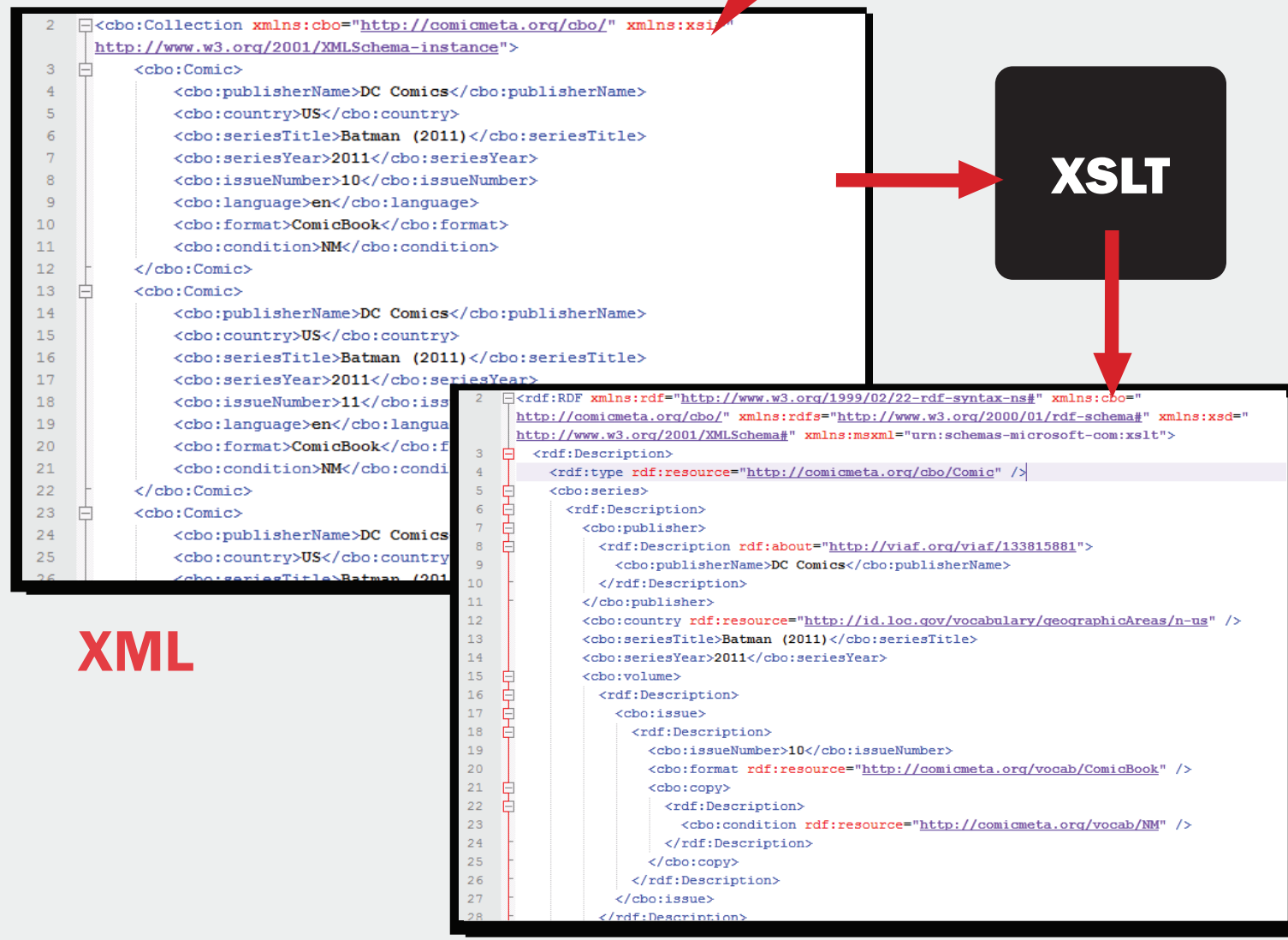
Users can prepare collection records by mapping spreadsheet (or exported CSV data) columns using an **XML map**. These records can then be exported directly from spreadsheet software, such as Excel (which makes this functionality available via the Developer toolbar). An alternative approach involves manually mapping columns by assigning new column headers that match element names in the core profile. The resulting mapped CSV data can then be uploaded to the **Core Data Converter** tool.

2. CONVERT TO RDF/XML

Once converted to XML, users can then generate RDF/XML records using the **Core Data Converter** tool, which uses an XSLT stylesheet for conversion. Common publishers, countries, and languages are converted to LD URIs, while popular comic book terms are converted to URIs in the **Comic Book Vocabulary**, a LD project being developed in parallel to the Ontology. The XSLT stylesheet approach allows users to apply the transformation locally using any scripting language, and modify or adapt it as necessary.



CSV/XML Map



FUTURE RESEARCH

The workflow above illustrates how comic book collection records can be converted to Linked Data using the **Comic Book Ontology**. Future research will explore extending this functionality to the broader comics community, and incorporating resources from existing projects like the **Grand Comics Database** (comics.org).

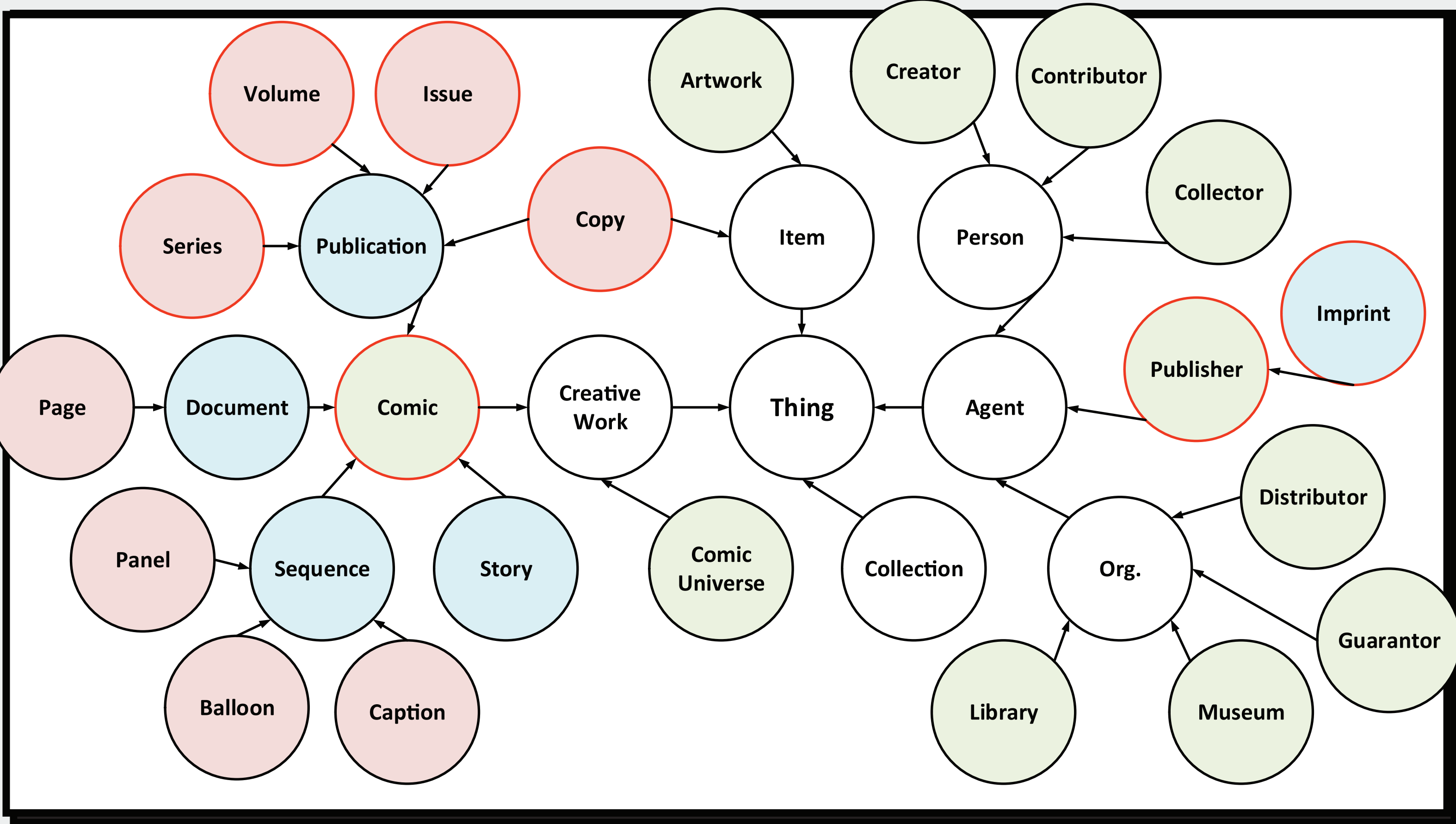


Fig. 1 - Comic Book Ontology (CBO) - <http://comickmeta.org/cbo>