

## Against the Grain

---

Volume 23 | Issue 1

Article 10

---

February 2011

# KBART -- Providing Standardized, Accurate and Timely Metadata: Methods and Challenges

Julie Zhu

*American Institute of Physics*, [jzhu@aip.org](mailto:jzhu@aip.org)

Gary Pollack

*Cengage Learning* | *Gale*, [gary.pollack@gale.com](mailto:gary.pollack@gale.com)

Ruth Wells

*Taylor & Francis*, [Ruth.Wells@tandf.co.uk](mailto:Ruth.Wells@tandf.co.uk)

Matthew Llewellyn

*Royal Society Publishing*, [Matthew.Llewellyn@royalsociety.org](mailto:Matthew.Llewellyn@royalsociety.org)

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

 Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Zhu, Julie; Pollack, Gary; Wells, Ruth; and Llewellyn, Matthew (2011) "KBART -- Providing Standardized, Accurate and Timely Metadata: Methods and Challenges," *Against the Grain*: Vol. 23: Iss. 1, Article 10.

DOI: <https://doi.org/10.7771/2380-176X.5733>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# KBART – Providing Standardized, Accurate and Timely Metadata: Methods and Challenges

by **Julie Zhu** (Sr. Project Coordinator, Online Service Division, American Institute of Physics) <jzhu@aip.org>

and **Gary Pollack** (VP Customer-Partner Solutions, Cengage Learning | Gale) <gary.pollack@gale.com>

and **Ruth Wells** (Journals Project Manager, IT Department, Taylor & Francis) <Ruth.Wells@tandf.co.uk>

and **Matthew Llewellyn** (eProjects Manager, Royal Society Publishing) <Matthew.Llewellyn@royalsociety.org>

**P**ublishers, librarians, and educators understand that metadata is an increasingly important aspect of resource discovery and use. We all know that good metadata or, better yet, standards-based metadata facilitates interoperability of services provided by our knowledge-base and learning management systems; ultimately connecting the communities of end users we serve to relevant and appropriate digital content.

In the age of mostly print publications, librarians were often responsible for creating cataloging and metadata information for journals and other publications subscribed by libraries. Now in the age of electronic publications, when more and more libraries are shifting to online-only subscription models and when many libraries are facing budget and staff shortages, libraries and library service providers are calling upon the content providers to provide publication metadata in a standardized, accurate, and timely way.

Several years ago some service providers and aggregators, like **Serials Solutions**, **Ex Libris**, **EBSCO**, started asking publishers and content hosting platforms to provide publications titles lists. **Serials Solutions** published a format for the metadata needed for serials and monographs, while other library service providers did not provide specifications. Some publishers started sending serials titles lists to these requesting library service providers via email, FTP, or Websites. The 16 standardized fields for serial titles specified in KBART Phase 1 Recommendations in many ways help the content providers, who do not have to modify the titles lists for different library service providers' knowledge bases.

While publishers fully understand the benefits of providing standardized, accurate, and timely metadata, they face practical challenges. Smaller publishers with only dozens of serial titles may produce and update their title lists in a manual or semi-automated process. This process will require designated human resources to periodically maintain and update the metadata. While many libraries are facing budget and staff shortages, publishers also experience staff shortages and competing projects. Providing metadata may not be at the top of some publishers' lists.

Larger publishers, hosting platforms, and aggregators cannot rely on manual or semi-automated processes. When hundreds or thousands of titles are involved, with backfile content sometimes added for some titles, and with frequent title changes, they have to use some automated processes. While they may

have more resources, they also face more competing projects and priorities. It is highly likely that the 16 required metadata fields are spread over multiple databases or systems, and it is also likely that metadata are not always accurate and up-to-date in these systems. To clean up legacy data and pull together metadata, just for serial titles, could become a major project for publishers.

What may not be obvious to librarians and educators are the vast sums of money and time that publishers must spend on systems with flexible metadata schemas, metadata schema views, metadata policies and processes, quality controls, collaborative metadata editing and authoring tools, and user-friendly interface components. While sometimes referred to as editorial workflow systems, these applications are increasingly being re-factored to deal with new requirements, whether internally driven or market driven, whether to meet a new or emerging standard or to accommodate a new type of digital asset (e.g., a "tweet"). In any case the system requires modification, and in order for that to take place requirements must be articulated, a project must be approved, a team must be formed, staff must be trained, etc.

The library community has raised more requests to publishers. Consortia would like to have serial titles customized for each consortium. Libraries would like to have metadata for monographs — i.e., online books and conference proceedings. Each request creates a new challenge for publishers. A publisher often serves dozens, and sometimes hundreds, of consortia. Even if publishers will only provide customized serial title lists for major consortia, they will need to maintain multiple lists and potentially increase the amount of work multifold. When multiple lists are maintained, there is always the possibility of them getting out of synch.

For most publishers, metadata for monographs and serials are stored in different ways. Therefore providing metadata for monographs could be a very different project. Since it is highly possible for publishers to have more monograph titles than serial titles and since monograph titles are added constantly, publishers may have to implement an automated process to generate monograph metadata.

Providing metadata for conference proceedings can be a more difficult challenge than providing metadata for online books. The first reason is that conference proceedings are a

hybrid of serial and monograph. The metadata should include information for the serial title as well as information on the volume level. A connection also needs to be made between the serial and the volume / monograph. The

second reason is that many conference proceedings include hundreds, or even thousands, of volumes and span decades. The quality of metadata for conference proceedings, especially for earlier volumes, can be quite poor. The third reason is since there have not been good standards for conference proceedings, the tagging of metadata for conference proceedings has been wildly inconsistent, for the same hosting platform, and sometimes even for the same publisher, or for the same proceedings across time. To provide correct metadata for conference proceedings, publishers must find ways to standardize metadata tagging and clean up legacy metadata.

As publisher systems and processes adapt to ever-changing market demands and as we increase the amount of metadata "attached" to an object, we do good things, including the likelihood of increased discoverability, but we also experience bad things, including taking on higher costs associated with producing metadata.

Despite the challenges, many publishers understand the ultimate benefits of quality metadata and are willing to make the commitment to provide improved metadata to the library community, not only for libraries and library service providers' knowledge bases, but also for major consortia, not only for serials, but also for monographs.

Some publishers have even moved beyond just managing serial and monographic publications into the realm of a vast array of digital assets and learning objects. In its simplest form, learning object metadata could be understood to be an electronic record containing data for a digital asset; much like a bibliographic reference card describes a book in a library. In more complicated terms, learning object metadata requires developing profiles to describe requirements (structural, semantic, and syntactic) and how they relate to workflow and storage. These more complex structures facilitate more intelligent relationships between the objects which allow for more intelligent connections in the Knowledge base and systems that support research and learning. 🌱

