



Electronic Resource Forum

Building the Global Open Knowledgebase (GOKb)



Kristen Wilson

NCSU Libraries, North Carolina State University, Raleigh, NC 27695, USA

ARTICLE INFO

Available online 18 November 2013

Keywords:

Electronic resources management
 Knowledge bases
 Open data
 Bill of materials

INTRODUCTION

The Global Open Knowledgebase project (GOKb) aims to facilitate the development of shared e-resources management data by developing a freely available, community-managed repository of key publication information about electronic resources as they are represented in the supply chain from content publishers to suppliers to libraries. This column will provide an overview of GOKb based on the author's experiences as a lead subject matter expert with the project, along with in-depth discussion of the GOKb data model, tools for data ingest and display, and plans for community management and sustainability.

© 2013 Elsevier Inc. All rights reserved.

1. What is GOKb?

At its core, the Global Open Knowledgebase project (GOKb) is a knowledgebase that will describe electronic resources, publisher packages, and platforms in a way that will be familiar to librarians who have worked with any of the existing commercial knowledgebases. Furthermore, GOKb will fill a new role in the knowledgebase ecosystem by enhancing the traditional model with several unique features. The first and most obvious is GOKb's community-managed, open data approach. The data found in GOKb will be made freely available to libraries, other knowledgebase providers, and the public using a Web-based interface and application programming interface (APIs). In addition to serving the community, open data offers practical benefits, including improved accuracy resulting from the expertise of many contributors and a dataset that is not tied to any one commercial product. The use of a community contributor model also means that users of the knowledgebase will have the opportunity to contribute to the project by helping to curate data, fix errors, and submit information about unique collections, such as regional offerings or consortial buying arrangements.

GOKb will offer a set of enhanced data elements, in addition to the basic information users expect to find in a knowledgebase. Like existing products, GOKb will offer lists of electronic book and journal titles grouped according to the many permutations in which publishers and vendors sell them — complete collections, subject collections, backfiles, and aggregated databases. For each of the titles it contains, GOKb will store Knowledge Bases and Related Tools (KBART)-compliant metadata such as standard identifiers, default coverage ranges, and URLs, in addition to fields that provide context, such as open access status or global availability. It will also include normalized organizations, which will be linked to the other entities in the system using a series of roles.

GOKb aims to go a step further by introducing several improvements to the standard set of knowledgebase metadata. The enhanced data model will track changes over time, including publisher take-overs and bibliographic changes. Information about changes will allow users to see not only a snapshot of the e-resources landscape at any given moment, but also to better understand the connections between resources and their historical and future contexts. GOKb will also be capable of storing an expanded set of identifiers, including digital object identifiers (DOIs), institutional identifiers like the International Standard Name Identifier (ISNI), and proprietary ID numbers from publishers and vendors who are willing to share them.

To understand exactly how GOKb will fit into the current knowledgebase environment, it is also important to talk about what GOKb is not. Unlike current knowledgebases, GOKb will not be a place to store information about any one library's individual holdings. Rather, it aims to provide an open data set covering all available offerings that can be used by the library community for reference, as a part of open source systems and projects or for any other purpose we can imagine.

The Kuali OLE project (<https://www.kuali.org/ole>) provides one example of how GOKb data might be used in an external system. Kuali OLE is a community-source library management system currently being developed under the umbrella of the Kuali Foundation, a group that coordinates the development of open source administrative software under the Educational Community License (<http://kuali.org>). The Kuali OLE software will take advantage of data provided by GOKb in two ways. First, GOKb will provide an initial data source that will allow libraries to create records for e-resources management within OLE. Rather than manually enter data into a series of records that will describe e-resource packages and titles, users will be able to search GOKb from within Kuali OLE, select the packages or titles they are purchasing, and import data into the system where it will populate records with management metadata. GOKb will also provide users of Kuali OLE with updates regarding changes to e-resources. If a journal changes its title or

E-mail address: kristen_wilson@ncsu.edu.

moves to a new publisher, for example, OLE users will be able to choose to receive alerts or set up automatic actions within the system to handle each situation.

Along these same lines, GOKb is not a knowledgebase primarily designed to support access; instead, it is aimed primarily at back-of-house management of e-resources. This means that, at present, GOKb will not include a link resolver, MARC record service, discovery layer, or other tools for directing patrons to e-resources. (Though, because the project provides open data, someone could certainly build one of those tools using GOKb as a data source.)

The benefits of a knowledgebase for e-resources management are already being seen by Jisc's Knowledge Base + (KB+) (<http://www.jisc-collections.ac.uk/knowledgebaseplus>). KB+ is a shared e-resources management tool created by Jisc (<http://www.jisc.ac.uk>), a UK-based non-profit that supports the development of technologies for higher education. KB+ allows institutions in the UK to access a centralized data store where they can find out about Jisc Collections negotiated packages, model licenses, and where they can upload their local subscriptions, holdings, entitlements, and license agreements. Many academic institutions in the UK negotiate their electronic resources purchases at the national level and manage many of the same resources, allowing users of KB+ to share the burden of collecting and maintaining this data. In building a shared knowledgebase, KB+ users have worked together to create a centralized source of reliable data that can be used to better understand the contents and terms of their collections. Because GOKb represents the logical extension of this concept from the national level to the global, KB+ has contributed its data to GOKb and will soon be more fully integrating its data and systems with the GOKb environment.

2. The GOKb Data Model

One of the primary goals of the first phase of GOKb development was to create a robust data model that can represent current e-resources purchasing arrangements and be extended to accommodate these arrangements as they evolve. To achieve these objectives, the GOKb development team has implemented a data model called the bill of materials (BOM), or parts explosion (http://en.wikipedia.org/wiki/Bill_of_materials), which provides the flexibility needed to manage current e-resource products at the appropriate level, track changes over time, and describe new publishing and sales models as they emerge.

The idea of using the BOM model to support e-resources was first introduced to the project by team members at Sero Consulting and Knowledge Integration, the project management and development firms for the project, respectively. It used fast food as an analogy for how the model would work. At McDonalds, for example, you can buy many different items grouped in many different ways. You can buy a hamburger or fries or a drink, but you can also buy all three together as a combo meal. There are also purchases that come with certain items that you can't buy separately – like pickles. McDonalds needs to manage its inventory of these components and the different combinations in which they can be sold. Thus, in the BOM model, individual items – like meat, bun, pickles, lettuce etc. – are labeled components. And the final product – the cheeseburger – is called a combination. A combination can even be part of a larger combination. That cheeseburger, for example, can be combined with fries, a drink, and a toy to make a Happy Meal. To put it simply, the bill of materials model allows for the creation of a list of raw materials needed to build an end product.

Electronic resources can be seen in much the same way. Journals are the most obvious components, but there are actually many other elements that make up the typical purchasing arrangement. When a library purchases the journal *Serials Review*, what it really purchases is a specific iteration of this journal. It could be the version hosted on the ScienceDirect platform, but it could also be the version hosted on the EBSCOhost platform – hence, the platform is another component of that purchase. Similarly, the package to which that title belongs is also

a component of the purchase. The title may belong to Elsevier's Freedom Collection or EBSCO's Academic Search Premier. To better describe the entity that a library purchases, the BOM model allows GOKb to form a new combination of these three elements: *Serials Review* (which in GOKb is called a title instance or TI), the Freedom Collection (a package), and ScienceDirect (a platform). This type of combination is known as the Title Instance Package Platform – or TIPP. The TIPP is an especially valuable concept because it models the entity that a library purchases and that the user accesses, unlike work- or manifestation-level representations such as the bibliographic record. Because of the linkages that are naturally part of the TIPP, users of GOKb can easily navigate from a TIPP to the Title Instance, Package or Platform to which it is linked and, from there, to other related titles. The TIPP can also be linked to other entities within GOKb, including organizations and standard identifiers.

In addition to describing an existing resource in the present, the BOM model will also help GOKb to manage changes over time. Because any two components can be used to form a new combination, titles, packages, and platforms can be linked together to represent the relationships created by changes in the supply chain. In one example, the journal *Disease Management* might change its name to *Population Health Management*. Each of those titles will be represented by a Title Instance component in GOKb and the two TIs will be linked through a TI-TI combo. Similarly, if the title *BioScience* has been owned by the University of California Press, but is sold to Oxford University Press, the Title Instance will be updated to include both publishers and associated dates recorded and linked to both the previous and the new TIPP. Creating a means of tracking the movements of titles between publishers should be especially valuable to librarians using GOKb for electronic resources management. The Quali OLE project plans to use GOKb to receive notifications of changes to titles contained in an institution's collection. When OLE gets an electronic notification that, for example, a title has been sold to a new publisher, users may opt to receive prompts to take certain actions, such as consulting the governing license agreement for implications, confirming the order with their subscription agent, checking to see they will maintain their spend with the old publisher, or activating the new version of the title in a vendor link resolver.

Another key benefit of the BOM model is that it will provide GOKb with the flexibility needed to model new publishing and sales models as they emerge. Because all entities within the data model are simply components or combinations, any unit of content sold by an academic publisher or vendor can be represented as one or the other. So if sales of individual journal articles begin to dominate the marketplace, GOKb could represent each article in a journal as a component, while the journal that contains them becomes a combination. Such extensibility may prove extremely valuable as models like pay-per-view and demand-driven acquisitions continue to gain traction among academic libraries.

3. Data Ingest and Maintenance

Even the best data model doesn't amount to anything unless it contains clean and useful data. One of the primary responsibilities of librarians working on the GOKb project will be to harvest data from publishers, standardize it, and maintain its accuracy as changes occur. To support data normalization and loading of new files over time with minimal redundant effort, the GOKb team adopted a rules-based approach.

The people loading data into GOKb and cleaning it up will be librarians who are experienced practitioners in the field of e-resources, but not necessarily programmers or technical experts. The goal of selecting a data cleanup tool for use during the initial ingest of data into GOKb was to provide a way for these domain experts to easily do what they know needs to be done to clean up data without having to focus on the small details of how to actually do it. Programmers from Knowledge Integration evaluated several open source rules engines and data

Download English Version:

<https://daneshyari.com/en/article/1099842>

Download Persian Version:

<https://daneshyari.com/article/1099842>

[Daneshyari.com](https://daneshyari.com)