# Maximizing the knowledge base: Knowledge Base+ and the Global Open Knowledgebase

*Based on a paper presented at the 36th UKSG Annual Conference, Bournemouth, April 2013*

The motivation for the two projects discussed in this article is the simple premise that the current inaccuracies of data in the library supply chain are detrimental to the user experience, limit the ability of institutions to effectively manage their collections and that resolving them is increasingly unsustainable at the institutional level. Two projects, Knowledge Base+ (KB+) in the UK and Global Open Knowledgebase (GOKb) in the USA, are working in cooperation with a range of other partners, and adopting a community-centric approach to address these issues and broaden the scope and utility of knowledge bases more generally. The belief is that only through collaboration at a wide range of levels and on a number of fronts can these challenges be overcome.

## Background

Currently, academic institutions across the globe are spending far too much time correcting and maintaining the knowledge bases upon which a variety of library systems and services are built. To make matters worse, much of this effort is duplicating work that is occurring elsewhere, a situation that would be frustrating in the best of times, but completely unacceptable and unsustainable in more austere times. Two projects, Knowledge Base+ (KB+) in the UK and Global Open Knowledgebase (GOKb) in the USA, are working in cooperation with a range of other partners, and adopting a community-centric approach to address these issues and broaden the scope and utility of knowledge bases more generally. The belief is that only through collaboration at a wide range of levels and on a number of fronts can these challenges be overcome.

LIAM EARNEY
Project Director
Shared Electronic Resource Management Support Service - Knowledge Base + (KB+)
Jisc Collections

## What are KB+ and GOKb?

KB+ is a new service from Jisc Collections to create a shared academic knowledge base for the UK academic community. It was funded via the Higher Education Funding Council for England's (HEFCE's) Universities Modernisation Fund in response to studies by Jisc and SCONUL that found significant demand from library directors for greater shared services in the areas of library management systems, e-resource management and licensing. KB+ wants to capture and represent information that institutions need to manage their subscribed resources, for example the details of titles in publisher packages, e-resource licences and institutional entitlement information. It is about providing a one-stop shop for the management of this information for and by UK institutions but also making sure that this information is widely available throughout the supply chain to any other systems, vendors and services that require it.

GOKb is a sister project of KB+ in the US. Funded by the Mellon Foundation, the aim of GOKb is to create a community-sourced knowledge base, primarily for use with Kuali's open library environment. Like KB+, with which it has worked on software development and data architecture, it intends to make all of the data it collects available under an open licence so it can be used in other systems. Whilst both projects have a shared interest in package and licence information, GOKb is less concerned with institutional-level data and more focused on the global-level information. As a result, the data being collected is complementary and the services function as sources of information for each other and their users.

## What is the problem that KB+ and GOKb are seeking to resolve?

"… the services function as sources of information for each other and their users."

The first problem is data quality and availability. It is no secret that there are a multitude of problems with the quality of data upon which so many library systems rely. Unfortunately, many publishers find it very difficult to provide accurate information on the titles they make available. Just to give some examples of this, KB+ has spent over 70 hours creating a title list for just one of the major STM publishers, which the publisher could not generate from its own system. Some publisher systems generate different title lists based on the department requesting the information, so the sales team has access to one list, whilst the back office staff have access to another list, neither of which can be reconciled with each other.

A related issue is the availability of data. Not all of the parts of the supply chain have access to all of the information that they need. Two particular examples of this are:

1) Subscription information: information that is important to ensure ongoing access, but which often goes missing as titles move from publisher to publisher and institutions store the information in a variety of digital and analogue ways.

2) Licence information: an entire class of information that is largely absent from the supply chain in any way that is meaningfully useful for systems, yet is one of the things Jisc Collections is most commonly asked about.

So there are gaps in knowledge of what is available, what has been bought and what one can do with it.

The next issue is duplication of effort. Marshall Breeding undertook a study in 2012[1] looking at and comparing the major knowledge bases. Breeding noted that the big four knowledge bases have about 80 FTE between them spent working on the data. Yet there were 'only minor points of differentiation in their comprehensiveness and quality'. As a result, one could suggest that there is substantial duplication of effort providing the same information across different systems and silos. Add to that the amount of time and effort which is going on at the institutional level to correct, add information and update knowledge bases, and you have a huge duplication of effort on the systems vendor side and on the library side, not just in the UK but globally. Now, that duplication of effort may have been tenable in other times, but certainly today with the range of pressures from tight budgets, to implementing open access policies and ensuring the best possible student experience, it is really not sustainable any longer to have academic librarians' time taken up working on maintaining and correcting what is essentially the same commodity information – that could be available to everybody just once, from one point of contact.

"… a huge duplication of effort on the systems vendor side and on the library side …"

Finally, there is interoperability of data. As anyone from an institution will know, there is a great number and variety of 'silos' to be managed, each giving a slightly different view of what they subscribe to and what they have access to. There are publishers, subscription agents, ERMs, link resolvers, filing cabinets, spreadsheets, library staff (who are often the only people in the institution who understand what has been purchased, why it was purchased, where all the licences are stored, etc.) and what consortia say an institution has access to. To build up the complete picture, there is a need to try and reconcile all of these different silos of information, which often do not communicate with each other very well, if at all.

## Maximizing the knowledge base

So what is meant by maximizing the knowledge base and what are the approaches to achieving this? This article will cover four aspects (though there are many others): open data, collaborative communities, enriched information and standards and best practice.

### Open data

Why is open data so important to these services? The interest stems not from an ideological view of the world. Rather, from a KB+ and GOKb perspective, openness of the data is the only thing that allows them to work in the way that they have to. It allows them to share and collaborate with everyone and anyone they need to. It can help improve accuracy by exposing the data to review; it means that publication information can be provided to and shared with any system, service, or vendor that users of KB+ and GOKb want and need it to be. Currently, OCLC, EBSCO, Serials Solutions and Ex Libris are all taking some of the data that KB+ makes available and using it to enhance their knowledge bases. In this way these services reduce the burden on any one part of the supply chain, by providing other sources of authoritative information.

"Open data reinforces the commitment not to be tied to any one system or service."

Open data reinforces the commitment not to be tied to any one system or service. So it does not matter, for example, if a library is a Serials Solutions or an Ex Libris customer.

### Collaborative communities

One of the common misconceptions about this work is that it is just about duplicating what is already there, with more accuracy. Whilst this is undoubtedly a large part of the story, it should also be noted that the services intend to achieve much more, through collaborative working at different local, regional, national and international levels. KB+ can contribute to and make use of globally relevant information held and managed through GOKb, with title information, publisher information, platform information and standard licence agreements. But there are important differences at the national and institutional level that are best managed closer to the point of need. Some examples of this might be details of national purchases of journal archives or e-resource agreements where the content licensed differs from what is globally available from a publisher. One might also add in specific national licences, such as the Jisc model licence, consortially negotiated entitlements that differ from the standard, or instances where the content is made available on different platforms in different territories so there are different URLs and access mechanisms.

Organizations such as Jisc Collections can help add important context to national-level information by providing the benefit of their direct experience and involvement in the licensing and the negotiation procurement process, but there is also the local institutional world of holdings, financial data and documentation that needs to be maintained.

Ultimately, this is an attempt to minimize the effort required for any local institutional data management by ensuring that if it can be done as well or better at the national or international level, then it is. And that the results of those efforts are shared appropriately with those third-party systems, services and vendors wherever and whenever they need it.

**International Cooperation** One of the main critiques of initiatives such as GOKb and KB+ is that, given the limited resources available to them, it will be impossible for them to offer the comprehensiveness of the existing major knowledge bases. And it is a fair point when you consider that as described earlier some of these services have maybe 30 full-time staff just working on the data. However, to a certain extent it also misses the point. At a national level there are already tens of librarians working by themselves or in groups aligned to products, to maintain and update knowledge bases. This effort could be coordinated, bringing benefits to all, reducing individual work, eliminating duplication of effort across institutions.

Similarly, one can collaborate internationally and maximize any network effects. Groups from the UK, France, Germany, Canada, Japan, Sweden, China and the USA, are already actively looking at how they can collaborate to share data, manage data, or make data available openly. Since KB+ and GOKb use open data and set no limits on what each of us can do with the data, they are free to go on establishing new partnerships with any group that wants to see this type of information openly available to all, irrespective of platform, system, service or vendor.

Not everyone needs to do everything. The work can be divided up to reflect different areas of expertise and different national priorities. But an essential part of the sustainability is based on the collective effort.

"… an essential part of the sustainability is based on the collective effort."

## Enriched information

It is becoming a common, if overly simplistic statement, that everyone wants data to work harder to deliver better results. From a UK perspective, there are a range of data services available to everyone[2] and all of which contain part of the overall picture about access, use, rights and management of resources. The objective, and the challenge, is to bring that data together in intelligent and appropriate ways that can provide answers that inform the decisions institutions need to take. For example, one might take holdings and licensing information from KB+ and bring it together with preservation information from the Keepers' Registry and the UK Research Reserve to inform decisions on print disposal.

Whilst the data sources themselves are not necessarily important, bringing multiple sources of information together and interrogating them in new ways means that one can learn more than was possible from each in isolation, saving institutions valuable time and effort in undertaking such investigations themselves.

**Human perspectives** Thus far the article has concentrated on the data, its management, improvement and collation but underpinning all of this is human interaction and relationships.

An excellent example of this is licensing. Anyone who has ever read a licence will know that what is written down on the page is often only half the story. Experience, knowledge of librarians, knowledge of the process of the licence are all needed in order to apply that information correctly and make sure that the maximum access is available to users. Far too often individuals within institutions are subjecting themselves to stress and worry because they don't know whom to ask or even what to ask.

There are still studies going on about walk-in access and whether or not it is permitted, despite the fact that few, if any, publishers forbid walk-in access and this has been the situation for a number of years. This suggests that the challenge facing institutions is about access to the information they require, understanding that information and confidence in the answers that one arrives at.

"… underpinning all of this is human interaction and relationships."

KB+ is putting in place mechanisms that will not only present the information in an easy-to-consume way, but will help maximize that collective knowledge within the community, making sure that the knowledge of people in organizations like Jisc Collections and the knowledge of librarians is shared, is made available, filling in the gaps, adding context, collectively responding to the common questions across the community.

## Standards and best practice

The final section is concerned with the role of standards and best practice in all of this. The use of standards and identifiers, and compliance with best practice guidelines, have always been central to both KB+'s and GOKb's concept of how to resolve the issues that have been discussed throughout this paper. Their implementation supports accuracy and availability in the exchange of data, which is what KB+ and Go KB are all about.

When the KB+ project started, the intention was very much that if the correct standard were implemented, one could minimize the need for any human manipulation of the data, creating nice, machine-based processes for the collation, maintenance and distribution of all of the data. However, there were two problems with this approach. The first was that frequently, even where a standard existed, it had not been implemented widely or indeed at all. And this was particularly the case with some of the ONIX standards, which are very rich but seldom made full use of.

The second problem was that often, even when they had been implemented, there were still ongoing problems with the accuracy and the quality of the data itself. It was not enough to just put data in the right format: it needed to be the right data and all of the data. So both services have adopted a practical approach. Where the standard or guidance has been implemented, it will be used, as in the case with KBART[3]. Sometimes KB+ and GOKb will put the data into the required standards themselves, again as in the case with KBART and things like ONIX Publishers Licence (ONIX PL). Other times identifiers will be mapped together so that one can build up an increasingly accurate picture of how the data fits together.

"It would be much more sensible if this work was undertaken at source …"

However, one questions whether this is the best way to proceed, whether it is appropriate for the academic community to do this and whether it is placing an unacceptable burden on the academic community to maintain the date of other organizations.

It would be much more sensible if this work was undertaken at source by the relevant publishers and suppliers who are publishing and are best placed to ensure the accuracy of the data.

The successful adoption of the COUNTER code for usage statistics demonstrates that where there is a will, these standards can be adopted. However, it also demonstrates the importance of developing a compelling business case for adoption of standards, and it would appear that thus far the library and library systems communities have not been able to do this for knowledge base information.

## Conclusion

KB+ and GOKb are community-sourced approaches to resolving long-term deficiencies in the accuracy and availability of knowledge base data in the library system and wider scholarly academic publishing supply chain.

"… where there is a will, these standards can be adopted."

By adopting a community-based approach centred around openness and collaboration, it is intended that these services can harness the efforts of the wider academic library community and add value that will not only reduce costs at the individual institutional level but also help them manage their collections more effectively.

References

1. Breeding, M, E-resource knowledge bases and link resolvers: an assessment of the current products and emerging trends, *Insights*, 2012, 25(2), 173–182; DOI: http://dx.doi.org/10.1629/2048-7754.25.2.173 (accessed 4 October 2013).
2. Examples could be JUSP (http://jusp.mimas.ac.uk/), SHERPA/RoMEO (http://www.sherpa.ac.uk/romeo/), KB+ (http://www.kbplus.ac.uk) and The Keepers Registry (http://thekeepers.org/thekeepers/keepers.asp).
3. Knowledge bases and related tools (KBART): http://www.uksg.org/kbart (accessed 7 October 2013).

Liam Earney, Project Director, Shared Electronic Resource Management Support Service - Knowledge Base+ (KB+)
Jisc Collections, Ground Floor, Brettenham House, 5 Lancaster Place, London WC2E 7EN
E-mail: l.earney@jisc-collections.ac.uk

To cite this article:

Earney, L, Maximizing the knowledge base: Knowledge Base+ and the Global Open Knowledgebase, *Insights*, 2013, 26(3), 244–249; DOI: http://dx.doi.org/10.1629/2048-7754.71