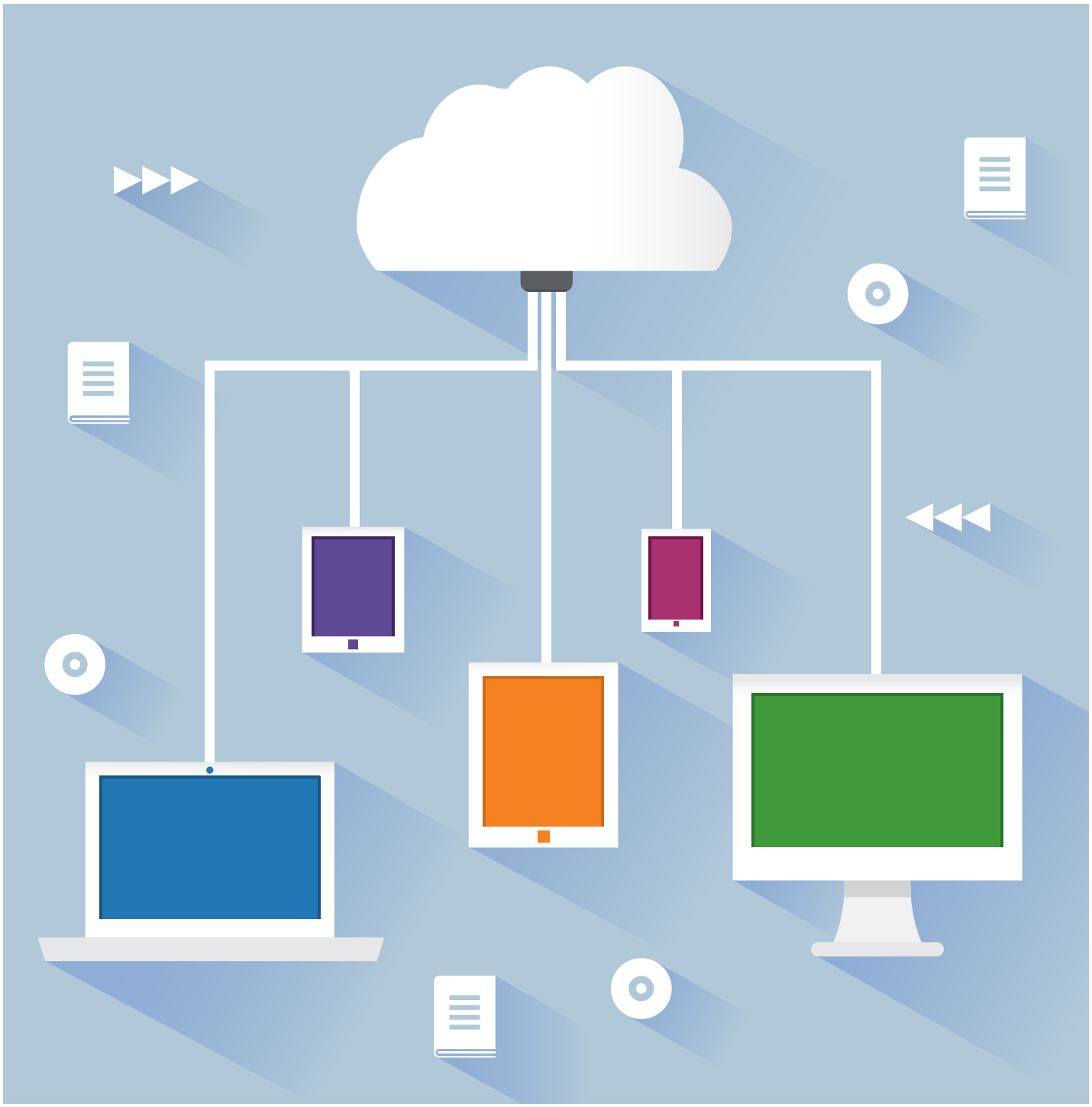
A CROSS-INDUSTRY WHITE PAPER

E-DATA QUALITY WORKING GROUP

Success Strategies for Electronic Content Discovery and Access



Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper
Suzanne Saskia Kemperman, Bill Brembeck, Elizabeth W. Brown, Alexandra de Lange-van Oosten, Theodore Fons, Catherine Giffi, Noah Levin, Alistair Morrison, Carlen Ruschoff, Gregg A. Silvis and Jabin White

September 2014

OCLC
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 1-55653-483-3 (electronic)
978-1-55653-483-6 (electronic)
1-55653-484-1 (print)
978-1-55653-484-3 (print)

OCLC Control Number #890816552 (electronic)
#890816452 (print)

Please direct correspondence to:
Suzanne Saskia Kemperman
Director, Business Development and Publisher Relations
OCLC
suzanne_kemperman@oclc.org

Suggested citation:
Kemperman, Suzanne Saskia, Bill Brembeck, Elizabeth W. Brown, Alexandra de Lange-van Oosten, Theodore Fons, Catherine Giffi, Noah Levin, Alistair Morrison, Carlen Ruschoff, Gregg A. Silvis, and Jabin White. 2014. *Success Strategies for Electronic Content Discovery and Access: A Cross-Industry White Paper*. Dublin, OH: OCLC.
http://www.oclc.org/content/dam/oclc/reports/data-quality/215233-SuccessStrategies.pdf.

*"Learn from yesterday, live for today, hope for tomorrow.
The important thing is not to stop questioning."*

–Albert Einstein

E-Data Quality Working Group Members

Suzanne Saskia Kemperman
OCLC, Director, Business Development and Publisher Relations

Bill Brembeck
OCLC, Senior Product Analyst, Data Services and WorldCat Quality Management

Elizabeth W. Brown
Project MUSE, Manager, Publisher Relations

Alexandra de Lange-van Oosten
Elsevier, Head of Third-Party Platform Relations

Theodore Fons
OCLC, Executive Director, Data Services and WorldCat Quality Management

Catherine Giffi
Wiley, Director, Strategic Market Analysis

Noah Levin
Springer Science+Business Media, Metadata Manager

Alistair Morrison
Elsevier, Senior Product Manager, ScienceDirect

Carlen Ruschoff
University of Maryland, Director, Technical Services and Strategic Initiatives

Gregg A. Silvis
University of Delaware Library, Associate University Librarian, Information Technology and Digital Initiatives

Jabin White
ITHAKA/JSTOR, Vice President, Content Management

Executive summary

Library users conduct an increasing number of information searches online without the help of a librarian. The library's discovery and access systems play an important role in helping users sift through and access the large amount of electronically published content.

When these systems have to rely on poor-quality data, users face a major barrier to discovery and access. When discovery and access fail, users get frustrated and librarians may choose not to renew subscriptions to content or to knowledge base and discovery services. A break in one part of the content supply chain will affect all links in the chain. However, by implementing some basic improvements, libraries, service providers and data suppliers (including, but not limited to, publishers) can remove this barrier and improve users' experiences.

The authors of this white paper, the E-Data Quality Working Group, are representatives of libraries, data suppliers and service providers. We recognize that all of us, as participants in the content supply chain, have a shared interest in improving content discovery and access for library users through better quality bibliographic metadata and holdings data. We also recognize that we have a shared responsibility to improve the quality of the data exchanged and to implement more effective data exchange workflows.

The content supply chain faces three core problems with the current state of data quality. All of these issues can prevent users from getting to the resources they need.

- **Data are incomplete or inaccurate.**
  This includes bibliographic metadata (needed for discovery) and holdings data (needed for access).
- **Bibliographic metadata and holdings data are not synchronized.**
  Libraries and service providers have difficulty maintaining knowledge bases when they receive data for a single item or collection at different times.
- **Libraries receive data in multiple formats.**
  Libraries must spend time and resources reformatting and completing the data, which introduces the possibility of localized error.

We identify recommended solutions to these problems with data quality. These recommendations should be implemented by all participants in the content supply chain to enhance resource discoverability and accessibility.

1. **Improve bibliographic metadata and holdings data.**
   - 1.1 **Use e-identifiers** instead of print identifiers in bibliographic metadata to describe e-resources.
   - 1.2 **Provide consistent collection information** to align data with the titles and collection names used in the sales and marketing materials.
   - 1.3 **Verify data before sending** to ensure that the data provided matches the library's actual holdings.
2. **Synchronize bibliographic metadata and holdings data.**
   - 2.1 **Follow a schedule** to update data files at the same time as collections.
3. **Use consistent data formats.**
   - 3.1 **Use Knowledge Bases And Related Tools (KBART) and Machine-Readable Cataloging (MARC) standards** to exchange data throughout the supply chain.
   - 3.2 **Provide change management records** with scheduled data feeds to alert libraries to alterations in collection subscriptions.
   - 3.3 **Provide direct holdings data to the service provider** so that libraries will no longer have to manage their holdings independently.

By working together to address cross-industry problems with data quality, parties involved in the content supply chain can improve the value of their content and their service to library users.

Background

One of the library's main missions is to get the right resources in front of its users at the right time. These days, the right resources often consist of databases, electronic books and electronic journal articles. Instead of maintaining all of these e-resources internally and separately, libraries purchase licenses for e-books, e-journals, databases and collections that are housed on data suppliers' websites. Library users discover the e-resource availability through one of many discovery tools—such as libraries' online catalogs or Google Scholar—and immediately access the full text.

However, through no fault of their own, library users sometimes find themselves unable to discover and access licensed e-resources provided for their use. Perhaps a student working on her thesis can't find recent journal articles related to her topic. Or a professor directs students to an e-book through the library only to discover that the link is broken. Why does this happen?

Library e-content is facing a data quality problem, which directly affects users' ability to find and—maybe more importantly—to use library resources. All parties involved in the content supply chain must put in place the structure and tools necessary to allow library users to efficiently access content at their points of need.

As the volume of papers published, data sets compiled and works cataloged continues to increase, finding the right piece of information is quickly becoming the search for a needle in a haystack. The key to sifting through this wealth of information is the background data that discovery and access systems use to identify resources and to make them available to library users. Library e-content is facing a data quality problem, which directly affects users' ability to find and—maybe more importantly—to use library resources. All parties involved in the content supply chain must put in place the structure and tools necessary to allow library users to efficiently access content at their points of need.

Following a recommendation from the 2007 United Kingdom Serials Group (UKSG) report, *Link Resolvers and the Serials Supply Chain*, UKSG and the National Information Standards Organization (NISO) jointly established the KBART Working Group to consider improvements to metadata exchange that will enhance content discovery and access. In January 2010, the KBART Working Group published *KBART Phase I Recommended Practice*. The group continued work on more specific improvements to metadata quality, and it published *KBART Phase II Recommended Practice* in March 2014. This second document reflects the deepening understanding of the importance of the library's knowledge base for discovery, access and management of online resources.

Building on this progress, we identified a need to articulate the problems caused by poor or incomplete data and the benefits of the efficient exchange of high-quality data among libraries, data suppliers and service providers. Our cross-industry E-Data Quality Working Group, which includes representatives from libraries, data suppliers and service providers who are also library users, convened to address this issue in June 2013. In addition, McGill University, the University of Toronto and the University of Maryland provided detailed descriptions of workflows for e-resources that include acquiring cataloging records and activating titles in various knowledge bases. Using these workflows, which revealed specific points of failure in library processes, and our own experiences with data creation and exchange, we jointly developed a set of practical recommendations.

As active participants in the content supply chain, we recognize that content discovery requires both rich bibliographic records and an accurate knowledge base of library holdings. Synchronization between the two is critical to enable users to discover and access content. This white paper combines business and practical information with best practices and recommendations for successful content discovery and access.

Why is this important?

The growing number of e-resources, including born-digital content, facilitates a transition to an online research environment. In this setting, users rarely benefit from the in-person help of a specialized research librarian, which makes user-initiated discovery and access more important than ever before.

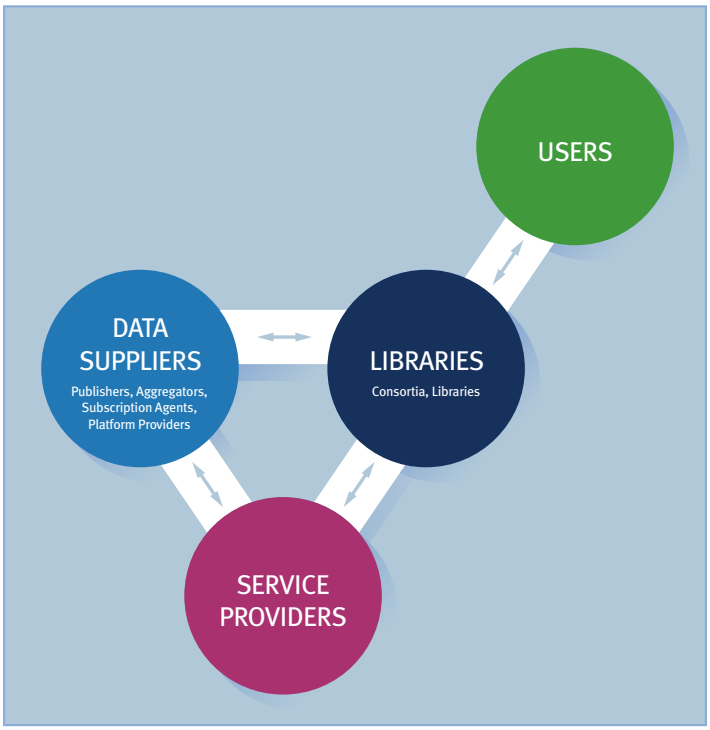


Figure 1.
There are many communication paths between content supply chain stakeholders. This paper focuses on the path from data supplier to service provider to library to user.

As figure 1 illustrates, there are many data exchange paths in the content supply chain, but this white paper focuses on a single workflow: the data exchange path from data supplier to library and user through the service provider. The ability to generate value for published content in libraries depends on data quality. Complete and accurate discovery depends on high-quality, rich bibliographic metadata. Selection of and access to full-text content depends on accurate holdings data for the specific library. Service providers or libraries enter this data into the library's knowledge base, which communicates with the library discovery service to provide users with access to e-resources. Without complete bibliographic metadata and holdings data, discovery and access can be frustrating or impossible to achieve.

Complete and accurate discovery depends on high-quality, rich bibliographic metadata. Selection and access to full-text content depends on accurate holdings data for the specific library.

Libraries, data suppliers and service providers have a shared interest in improving the flows of bibliographic metadata and holdings data that enable users to discover and access content. When one segment of the content supply chain faces a gap, it constrains the performance and business outcomes of all segments.

Users

Improved data quality benefits library end users most of all, as it enables them to efficiently discover and access information at their points of need. These users include researchers, faculty members, students and other information seekers. They should not miss out on relevant content because of poor bibliographic metadata or inaccurate holdings data, and their efforts to access content should not lead to a dead end.

Libraries

Improved data quality and automation allows libraries and consortia to more quickly provide purchased and licensed items to library users, which lowers the cost per use of each e-resource and increases its impact. Recent assessments at the University of Maryland indicate that the library requires as many as three full-time equivalents to maintain the knowledge base with the current workflows. Standardizing and automating the flow of bibliographic metadata and library holdings information allow staff members to focus on providing other important services to users rather than working on tasks that computers can easily do, such as downloading, reformatting, auditing and configuring data from disparate sources. It also frees up library budgets to shift from administrative expenses to more content purchases or other value-added activities. Data quality improvements will ensure that the money and staff time invested in managing libraries' online holdings and systems result in the greatest possible benefits for the users they support.

Data suppliers

Although many data suppliers are publishers, this category also includes content aggregators, platform providers, subscription agents and others who provide data and who would benefit from improved data quality processes. Without accurate data, overall usage of data suppliers' content is fundamentally low because users cannot find or access it. Effective discovery and easy access drive usage, which increases the value of the content to libraries and makes it more likely that the library will renew its subscriptions or will purchase new content.

Effective discovery and easy access drive usage, which increases the value of the content to libraries and makes it more likely that the library will renew its subscriptions or will purchase new content.

Libraries increasingly make decisions to purchase content or to continue subscriptions based on measurable, demonstrated value, such as usage statistics. When libraries analyze the value of their e-resource purchases and their return on investment, they look at the number of times a full-text title has been accessed to determine its cost per use. If access is low, cost per use is high. Both measures depend on the user being able to move from the discovery record to the full text without a hitch.

Service providers

Providers of discovery, article linker, electronic resource management (ERM), integrated library and other systems meet the needs of libraries and their users by transforming data from hundreds of data suppliers into a coherent system. Standardizing and automating the flow of content metadata and library holdings information will save service providers from interpreting, normalizing and re-presenting information from each data supplier. Improvements to the data quality workflow will reduce the amount of time service providers spend acting as interpreters and intermediaries between libraries and data suppliers. It will also increase service providers' ability to focus on the primary services they provide for libraries and users.

Problem statement

The number of electronic books, articles, journals and collections available through the library continues to grow, and users need to be able to find and use these resources. We have identified three key data quality problems that impede discovery and access, as shown in figure 2. Users will be unable to find the information they need if libraries are unable to maintain an up-to-date knowledge base with complete bibliographic metadata and accurate holdings data. Libraries, data suppliers and service providers will face increased costs to resolve these issues the longer they persist.

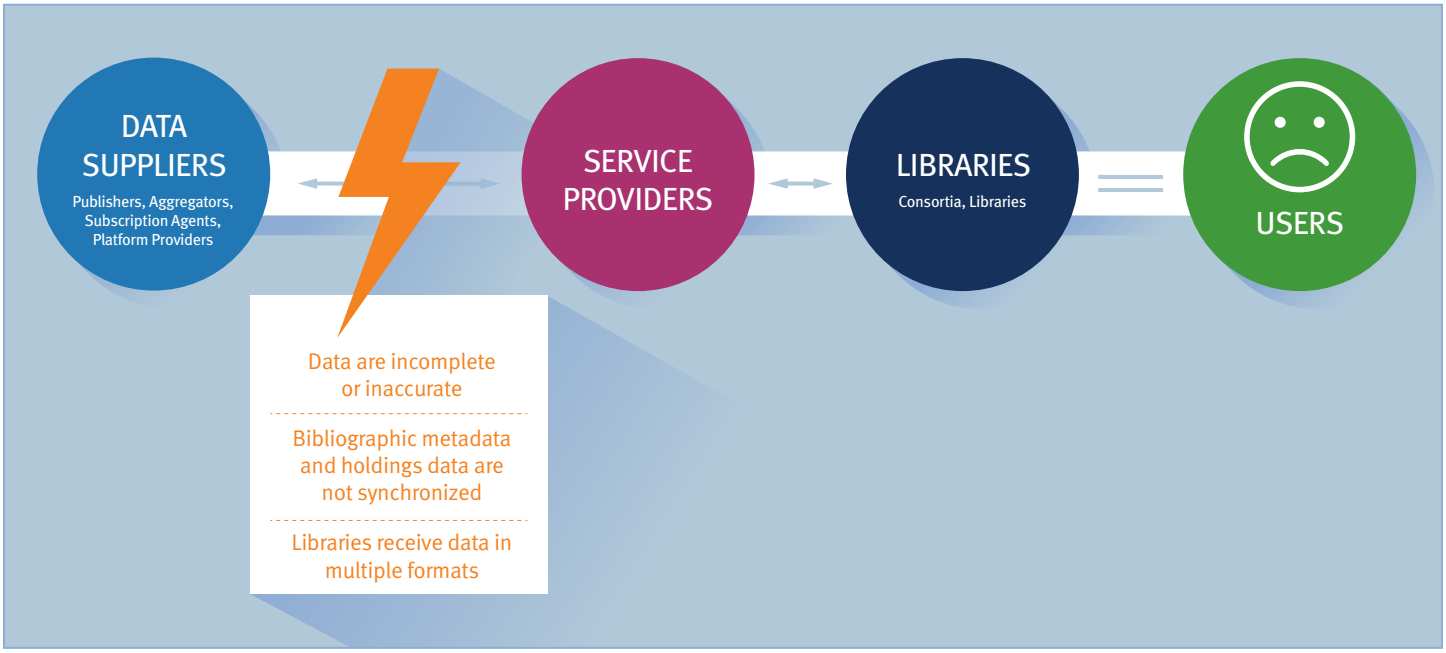


Figure 2.
Three main problems impede discovery and access of library e-resources.

Problem 1. Data are incomplete or inaccurate.

Libraries, data suppliers and service providers expend considerable resources to facilitate discovery and access for library users. Without accurate bibliographic data, users cannot find the content they need; without accurate holdings data, users cannot determine whether the resource is available to them.

There are multiple areas in which data could be incomplete or inaccurate.

- **Bibliographic metadata**
  Catalogs, ERM systems, discovery services, article linkers and other systems rely on complete and accurate bibliographic metadata in KBART and MARC files to discover and provide access to an item. Often, data suppliers find it easier to provide print identifiers for e-resources, even though print identifiers do not include of all the information needed for a complete bibliographic metadata record of an e-resource. Commonly, incorrect title information or incorrect coverage data (including unclear embargo information) can prevent users from discovering and accessing resources that are available to them.

- **Holdings data by item**
  Even when correct bibliographic metadata is available, the lack of accurate holdings data will lead discovery and access services to exclude or treat as unavailable items that should be included in the library's knowledge base. If each library does not correctly indicate its current holdings, users—including other libraries—cannot know what resources are available.

- **Holdings data by collection**
  Libraries frequently purchase collections or packages of e-content from data suppliers. However, sometimes the data supplier provides bibliographic metadata for items that differ from the actual content the library receives. If the library fails to correct this problem, the knowledge base will contain inaccurate holdings information. Further, depending on the library to make these corrections introduces the possibility of errors in the bibliographic metadata and variation between libraries. Another problem with collections holdings data is that knowledge bases often include only the most frequently requested packages. Data suppliers regularly update and modify collections and create new ones, which makes it difficult for librarians to keep track of what they're receiving and to maintain this information in the knowledge base.

Problem 2. Bibliographic metadata and holdings data are not synchronized.

Librarians need both rich bibliographic metadata and holdings data to maintain an accurate and useful knowledge base. Data suppliers and service providers do not always provide bibliographic metadata and holdings data simultaneously. When libraries receive the necessary data in separate parts, users may encounter a bibliographic record through the library discovery service that has been removed from the knowledge base. Alternatively, the user could have access to items in the knowledge base that are not discoverable through the library discovery service.

Problem 3. Libraries receive data in multiple formats.

Not all data suppliers adhere to a standardized, agreed-upon format when providing bibliographic metadata and holdings data to libraries. The data that a library receives from one data supplier may be in a different file format or may include different information than data from another data supplier, and libraries typically receive data from many suppliers. The bibliographic metadata that libraries receive may include inconsistent author information, subject headings or identifying numbers (e.g., ISBN, ISSN, PubMedID, OAI, CODEN and DOI). Libraries then have to spend time reformatting this data and filling in as many gaps as possible to ensure that the knowledge base has the most correct and up-to-date information available. Table 1 provides examples of the variation that exists among data suppliers' data as of August 2014.

Table 1.
Variations in data format and delivery methodology among data suppliers

Information accurate as of August 2014.

Bibliographic metadata								Holdings data		
Data supplier	Data exchange model	Data delivery method	Format	Frequency	Includes e-books?	Includes journals?	Provides update manifest?	Provides direct holdings?	Holdings level	Institution identifier mapping
EBL	Item	FTP	KBART	Weekly	Yes	No	N/A	Yes	Item	Provider-specific
Springer	Collection	Static link page	KBART II	Monthly	Yes	Yes	No	Yes (library-initiated)	Item	Provider-specific
Elsevier	Item	API and static link page	KBART II	Dynamic	Yes	Yes	N/A	Yes (via API)	Item	Provider-specific
	Collection	Static link page	KBART II	Monthly	Yes	Yes	No	No	Collection	Provider-specific
Sage	Collection	FTP	KBART II	Weekly	Yes	Yes	No	No	Collection	Provider-specific (Ringgold internally)
JSTOR	Collection/ Item (books)	FTP	KBART	Weekly	Yes	Yes	No	Yes	Collection	Provider-specific
Wiley	Collection	FTP	KBART	Monthly	Yes	Yes	No	No (planned)	Collection	Provider-specific

Recommendations

To support improved discovery of and access to full-text content for library users, our cross-industry working group developed a set of core recommendations related to data quality, as depicted in figure 3. Resolution of the issues outlined in the Problem statement section will require the coordinated efforts of data suppliers, libraries and service providers—the key stakeholders in the content supply chain and data management ecosystem. By adopting these solutions, we can jointly enhance service to information seekers and improve the value of electronic content in libraries.

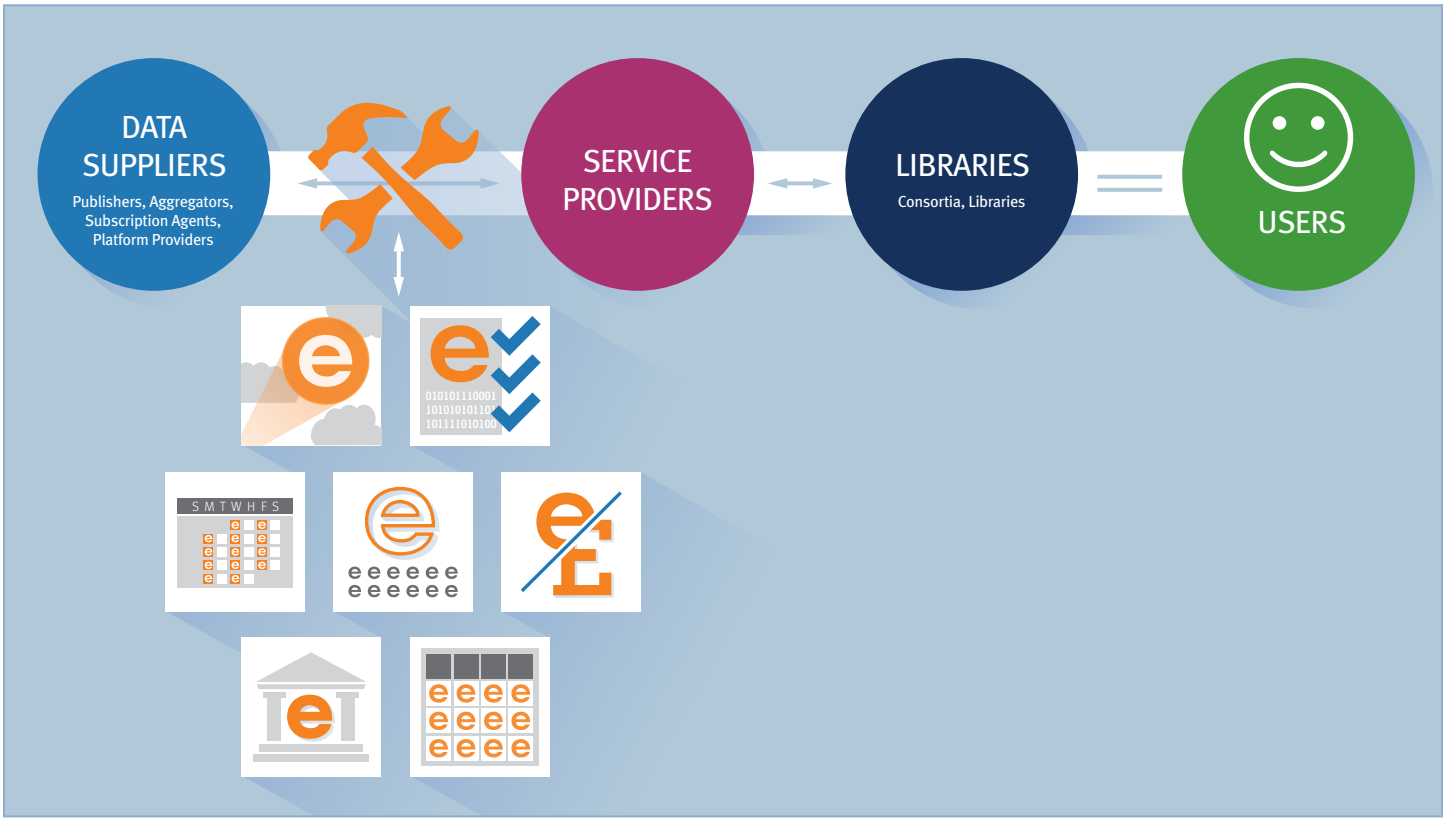
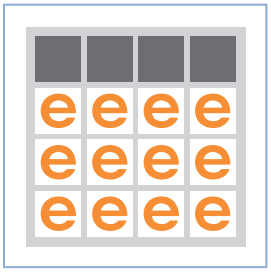


Figure 3.
Some basic improvements, if applied by all parties, can greatly enhance e-resource discovery, access and usage.

Recommendation 1:
Improve bibliographic metadata and holdings data.

Recommendation 1.1: Use e-identifiers.

Data suppliers should create and provide identifiers specifically for the e-resources in their collections instead of distributing print identifiers that lack important bibliographic metadata and result in false matches and false drops. By using only print identifiers, data suppliers avoid the cost of creating separate e-identifiers but contribute to the data quality problem.



Recommendation 1.2: Provide consistent collection information.

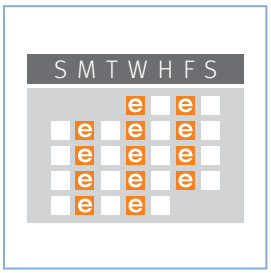
Especially when describing active and archival collections, data suppliers should use definitions, content descriptions and identifiers that the library will recognize. Likewise, service providers should not adjust collection names that they receive correctly from data suppliers. When librarians can match a collection to the same name and content descriptions they saw in the sales materials, their confidence in the data supplier and the collection's contents will increase. Additionally, libraries will spend less time conducting local reconciliation.

Providing consistent collection information is the best practice for archival collections that contain content that the provider no longer actively markets. It is critical for the maintenance of collection descriptions and holdings to include archival content with consistent content descriptions.

Recommendation 1.3: Verify data before sending.

Data suppliers should ensure that the data they send to service providers reflects what is available to the library. For example, if an item is removed from a collection but is still available to libraries that subscribed before the change, then data suppliers should ensure that those items remain in the holdings data with accurate linking information. Without this, users will not be able to discover or access the item through the knowledge base, even if it should be available to them. Likewise, if an item is no longer available through the library's subscription, then data suppliers should adjust the holdings data for this item to keep the knowledge base accurate.

Recommendation 2:
Synchronize bibliographic metadata and holdings data.



Recommendation 2.1: Follow a schedule.

While not all collections are updated at the same rate, data suppliers should provide bibliographic metadata and holdings data on a regular basis in a synchronized manner. The frequency of the data update should be consistent with the frequency of the collection update. If the data supplier updates a collection on a daily basis, then the data supplier should provide daily data feeds; if the data supplier updates a collection on a weekly basis, then the data supplier should provide weekly data feeds. This regular schedule will help libraries anticipate when new data will arrive, will improve the accuracy of the data in the knowledge base and will allow libraries and service providers to identify and correct problems in a timely manner to minimize the impact on users. Data suppliers should maintain open communication channels with service providers and libraries to provide timely information about schedule changes. In addition, service providers should load the received data promptly.

Recommendation 3:
Use consistent data formats.

Recommendation 3.1: Use KBART and MARC standards.

Libraries, data suppliers and service providers should use common, standardized data exchange formats, such as KBART and MARC, to reduce costs for all stakeholders. If all parties use these well-documented standards, the individual data elements will be better defined and less ambiguous. Stakeholders can communicate more clearly when they use the same terms, making troubleshooting much less complicated and less time consuming. Libraries will spend less time formatting and editing bibliographic metadata and holdings data, and service providers will spend fewer resources developing systems that compile disparate data into a single knowledge base or discovery system. Section 6 of *KBART Phase II Recommended Practice* offers detailed specifications for exchanging data with knowledge bases, which data suppliers should incorporate into their standard practices.

Recommendation 3.2: Provide change management records.

Data suppliers should provide details about changes to items and collections in their regular data feeds. Clearly alerting the service provider and the library about a change gives them an opportunity to verify that the change is made to the knowledge base, improving the discoverability and accessibility of that item. This clear acknowledgment also strengthens confidence in the accuracy of the knowledge base. See the Change management section for further guidelines on providing a collections update manifest.

Recommendation 3.3: Provide direct holdings data to the service provider.

In addition to bibliographic metadata and collection information, the data supplier should deliver direct holdings data to the service provider (and, when the library purchases and receives data directly from the publisher, to the library). The data supplier should indicate not only collection definitions but also which titles each library holds, an extension of KBART that allows records to be specific about each library's holdings.

When data suppliers offer service providers direct holdings data for mutual customers, they speed up access for library users and help libraries lower the operational costs of their staffs. With this information, the service provider can automate the process of setting holdings in the library's knowledge base, which eliminates the library's manual effort of adding holdings and increases the accuracy of the knowledge base holdings data. Less effort for librarians results in better, faster service for information seekers. See the Direct holdings section for additional details on suggested direct holdings information and the importance of site identifiers.

Implementation of recommendations

For discovery and access of library items to improve for users, stakeholders must not only agree to the recommendations above but must also agree on the data format and workflows for implementing them. Figure 4 illustrates an enhanced workflow that incorporates the recommendations.

Although maintaining this data can be challenging, these suggestions offer a way to overcome some of the initial barriers to the more technical aspects of our recommendations. We also encourage stakeholders to adhere to the detailed specifications for exchanging data with knowledge bases explained in Section 6 of *KBART Phase II Recommended Practice*.

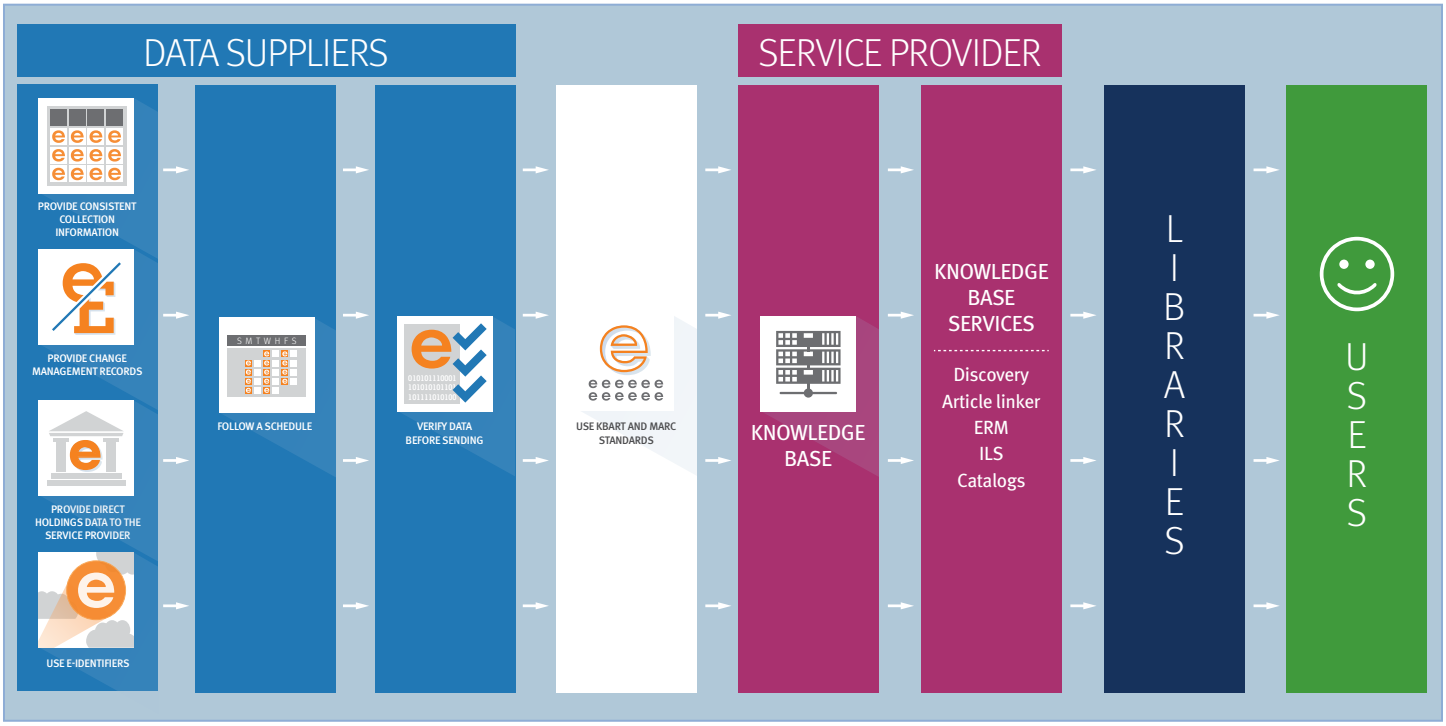


Figure 4.
Data quality can be improved only if all stakeholders commit to an optimal workflow.

Change management

Data suppliers make both scheduled and unscheduled changes to collections during the period of subscription. For example, libraries can purchase collections with e-books that are not yet published; once the e-books are published, the library must add their data to the knowledge base. If the data supplier fails to alert the library or the service provider that an e-book is available, or if the library or service provider does not update the knowledge base, then users will be unable to discover and access it even though it should be available to them.

When data suppliers provide collections data, the collections are not always recognizable to librarians working with knowledge base data. The collections data should be formatted to match the data supplier's current offering of subscription and access models, which will assist libraries and service providers in maintaining the correct holdings data.

We recommend that data suppliers provide change data through standardized update manifests of collections information. The library and service providers can use this manifest to accurately track changes and updates in the knowledge base. Table 2 offers some suggested information to be included in the collections update manifest. If data suppliers agree to use the same format for this information, libraries and service providers will expend fewer resources incorporating it into the knowledge base.

Table 2.
Recommended information to include in a collections update manifest

Suggested data	Description
Collection name	Human-readable name for this collection that aligns with the name used in the data supplier's marketing and sales materials
Collection identifier	Identifier used for this collection by the data supplier's system (sometimes called a database identifier), which should remain static and free of punctuation
Selectable	Yes/no indication of whether a library can select individual records from this collection through a subscription to a greater collection
Custom coverage	Yes/no indication of whether the data supplier will modify coverage dates for individual journals on a per-library basis
Open access	Yes/no indication of whether this collection contains only open-access titles
Patron-driven acquisition	Yes/no indication of whether this collection contains only titles available for patron-driven acquisitions
Collection URL	A URL that leads to a description of the collection (such as a search form or a listing of the titles in the collection) on the data supplier's website
Record count	The number of records in the collection, which is used to help verify the contents of the data files

Direct holdings

A knowledge base uses global subscription-level collection and title lists as a foundation, but having holdings for each library makes the system most meaningful to users and streamlines e-resource management for the library. Traditionally, each subscribing library has managed this data manually; service providers have only needed to know the data supplier's basic subscription offerings. However, knowledge bases would be easier for libraries to manage if the service providers also receive a full representation of the global holdings data and incorporate this information into their systems.

By providing direct holdings feeds with regular metadata updates, a data supplier can significantly ease the burden of maintenance and increase the satisfaction of libraries with their products. Table 3 outlines suggested information for all data suppliers to include when providing direct holdings data. The site identifier is discussed further in the next section.

By providing direct holdings feeds with regular metadata updates, a data supplier can significantly ease the burden of maintenance and increase the satisfaction of libraries with their products.

Table 3.
Recommended information to include in library direct holdings data

Suggested data	Description
Site identifier	A site or customer identifier
Collection identifier	An identifier that matches the collection identifier value in the collections information (not required for title-only holdings)
Title identifier	An identifier that matches one of the title identifiers in the data file for this collection (not required if the library subscribes to the entire collection without overrides)
Date of first issue online[1]	Date of the first issue that is available online; for e-books, the date of publication[2]
Number of first volume online[1]	Number of the first volume that is available online (not required for e-books)
Number of first issue online[1]	Number of the first issue that is available online (not required for e-books)
Date of last issue online[1]	Date of the last issue that is available online[2] (not required for e-books or if coverage is provided through the present)
Number of last volume online[1]	Number of the last volume that is available online (not required for e-books or if coverage is provided through the present)
Number of last issue online[1]	Number of the last issue that is available online (not required for e-books or if coverage is provided through the present)
Title URL[1]	Title-level URL (if different from the global URL)
Status	Optional information for sending changes only (not full files) to indicate whether this is a new holding (ADD), an existing holding to be removed (DELETE) or a changed holding (UPDATE)

1. This information is required only if the library's subscription is different from the data in the global holdings data files.
2. All dates should be presented in International Organization for Standardization (ISO) format: YYYY-MM-DD or YYYY-MM or YYYY.

Site identifiers

Although site identifiers are essential for direct holdings data, there is no universal site identifier authority. Data suppliers, service providers and libraries must map site identifiers between their systems to maintain useful data in the knowledge base. Data suppliers sometimes require site identifiers (including account or custom location identifiers) to access linked content, especially when libraries may individually configure subscriptions.

Until the development of universal site identifiers, data suppliers, libraries and service providers should map their specific site identifiers to ensure that the knowledge base contains only correct and useful site identifiers. When the site identifiers map between parties, users are significantly more likely to be able to efficiently discover and access the resources that are available to them.

Table 4 offers recommendations on how stakeholders can map their site identifiers to ensure that the knowledge base includes the best direct holdings information. When the library is part of a consortium, all parties must also map parent-child relationships between the libraries and the group.

Table 4.
Recommended information to include in a site identifier mapping file

Suggested data	Description
Site identifier	The site or customer identifier that matches the value in the holdings data
Site name	The name of the site or customer
Institutional symbol	The institution's symbol used by the service provider (multiple symbols should be separated with a comma, e.g., "TRN,SER,OZY")
Local site identifier	The identifier for this site or customer used by the data supplier's system
Country	The country where this site is located
Notes	Any other pertinent information
IP range	The data supplier IP address access range, if applicable

Conclusion

If libraries, data suppliers and service providers work together to improve and synchronize bibliographic metadata and holdings data and to use consistent data formats, then they will alleviate many of the challenges associated with users' discovery of and access to e-content. These stakeholders have a shared interest in improving the flows of data that enable library users to discover and access content, which will only become more difficult the longer the existing data workflows persist. The ability to generate value for published content in libraries depends on improved data quality and workflows. It is therefore of paramount importance to quickly implement improvements that allow users to efficiently access content at the point of need, especially as online, user-initiated discovery of e-resources becomes the primary means of conducting research.

If libraries, data suppliers and service providers work together to improve and synchronize bibliographic metadata and holdings data and to use consistent data formats, then they will alleviate many of the challenges associated with users' discovery of and access to content.

This white paper only begins to address the problems facing data quality. We focused on the biggest and most important issues, and we hope that this work will be endorsed and carried on by others. User discovery and access would be greatly enhanced if additional standards and guidelines existed to help libraries, data suppliers and service providers manage their bibliographic metadata and holdings data. In addition, stakeholders should consider developing universal site identifiers to ease the management of holdings data.

Our cross-industry group has outlined the first steps toward data quality improvement with the goal of making it easier for users to find and access library resources. The next steps require all parties involved in the data exchange process to make urgent changes to their existing data formats and workflows. Only by working together through all levels of the content supply chain can we ensure that users have the tools and opportunities required to find the materials they need.

Recommendations

- Use e-identifiers.
- Provide consistent collection information.
- Verify data before sending.
- Follow a schedule.
- Use KBART and MARC standards.
- Provide change management records.
- Provide direct holdings data to the service provider.

For more information

Culling, James. *Link Resolvers and the Serials Supply Chain*. Oxford: United Kingdom Serials Group, 2007. http://www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf.

KBART Phase II Working Group. *KBART Phase II Recommended Practice*. Baltimore: National Information Standards Organization, 2014. http://www.niso.org/apps/group_public/download.php/12720/rp-9-2014_KBART.pdf.

KBART Registry. https://sites.google.com/site/kbartregistry/.

National Information Standards Organization (NISO). http://www.niso.org/home/.

NISO/UKSG KBART Working Group. *KBART Phase I Recommended Practice*. Baltimore: National Information Standards Organization, 2010. http://www.uksg.org/sites/uksg.org/files/KBART_Phase_I_Recommended_Practice.pdf.

NISO/UKSG Knowledge Bases And Related Tools (KBART) Working Group. http://www.niso.org/apps/group_public/workgroup.php?wg_abbrev=kbart and http://www.uksg.org/kbart.

United Kingdom Serials Group (UKSG). http://www.uksg.org/.

ISBN: Electronic 1-55653-483-3 978-1-55653-483-6
Print 1-55653-484-1 978-1-55653-484-3
1409/215233, OCLC