What is all this fuss about? Is wrong metadata *really* bad for libraries and their end-users?

This article has been developed from our UKSG Conference breakout session 'What is all this fuss about? Is wrong metadata really bad for libraries and their end-users?' held in Glasgow in April 2018. The aim of the session was to develop an understanding of the issues caused by poor quality metadata in library workflows. We showed how poor metadata affects libraries' end-users despite the efforts of librarians to provide them with the best service by investing in expensive content and discovery systems. We highlighted the main challenges, why this matters and the impact that inadequate metadata is having, and concluded with some recommendations for stakeholders. In this session we asked the audience to write down and share with us similar issues concerning or affecting them, and some of these anonymous anecdotes are shared throughout the article. The complete responses can be found in the appendix at the end.

Keywords

Metadata; end-users; resource discovery; value; discoverability; electronic resources

Introduction

Metadata is beginning to seem like the new catch-all term for technological issues facing us all in an era preoccupied by uses of personal information, applications applied to big data, and data breaches. But, while we might live in unprecedented times in terms of the scale and function of dynamic metadata, the need for metadata itself, of course, is not 'new'.[2]

Metadata has always been at the heart of library services because they need it to describe their resources for end-user discovery and collection management. Without metadata, a room full of books is just a room full of books.

There are many definitions of metadata. The basic definition is 'data about data', which highlights the descriptive element of the term and its usage.[3] Metadata supports several library functions (resource acquisition and activation, resource discovery and collection management) and its creation, gathering and sharing are governed by principles, techniques, formats, and standards that are maintained and reviewed by international bodies such as the Library of Congress, the RDA Steering Committee, IFLA (the International Federation of Library Associations and Institutions) and NISO (the National Information Standards Organization). These standards and developments are then filtered back to librarians via national libraries and professional organizations. Metadata quality depends upon its purpose and the context of its creation and usage but, generally, librarians are looking for completeness, accuracy and timeliness of delivery when assessing metadata quality.[4] With e-resources increasingly dominating libraries' collections, the role of metadata in libraries has come under the microscope recently but it has not always been well understood or well managed. Despite its importance, for a long time metadata for e-resources has been treated as a luxury accessory rather than as something essential to resource discovery and access.

As methods for storing metadata electronically have developed, new opportunities for sharing it have inevitably emerged: with other libraries, with aggregators, with systems and so on. It is well known that the performance of library applications and content platforms is

MAGALY BASCONES

Metadata and Discovery Manager Digital Resources Division Bloomsbury Publishing Plc[1]

AMY STANIFORTH

Resource Discovery Team Leader Aberystwyth University

based on the accuracy, completeness and timeliness of the metadata that circulates within the supply chain – from creation to delivery.[5]

One UKSG participant suggested this may not always work as intended:

> 'Using Syndetics Unbound – pulls in some strange information due to poor metadata in our record, e.g. ISBN for a book in an e-journal catalogue record pulled in lots of unexpected data.'

And while new metadata requirements for e-resources have been developed (to describe new formats, access requirements, licensing and article- and collection-level description), the metadata that libraries and end-users require to discover and access resources has not actually changed much. They really do still need creator information, titles and dates of publication, for example, as well as subject headings and summaries. These are necessary both for the end-user *to find, identify, select* and *obtain* (FISO)[6] resources and, increasingly, for disambiguation in resource management.

However, due to poor quality metadata supply combined with complex and continuous processing, librarians rarely know the status of their library holdings and collections of e-resources. Librarians, publishers and system vendors are only just beginning to admit to the consequences of relying on e-resource metadata that, unlike that in a card catalogue, is constantly being put through processes of maintenance, merging and migration. Libraries are increasingly linking to metadata stored centrally in national or commercial knowledge bases where updates to vast electronic packages can be made in bulk, and ideally once, but these happen outside the control of the library.

On the one hand, this access method has meant that libraries have seen an increase in the volume of resources that end-users, theoretically at least, have access to. But on the other hand, it has also meant that libraries have assumed access and discovery of these resources to be straightforward. The solutions promised by system vendors reinforce this belief, they promise 'easy access', 'automation', 'interoperability', 'superior experience', 'customization' and 'streamlined processes'.

Yet, librarians today have become receivers of an uncontrollable flow of not always useful information. Libraries have adapted their workflows to integrate the proposed automatic solutions without critically assessing the consequences, and over time have reduced their cataloguing and metadata capacity and expertise so that they are less equipped as institutions to successfully identify and source the metadata needed for end-users to access library resources.

The problems described in this article originate from the practical realities of the academic publishing supply chain. Librarians are part of this chain, and they struggle between the constraints of trying to ensure end-user satisfaction and value for money and the aspiration to develop successful end-to-end workflows from acquisition to discovery of e-resources that automate tasks and release staff time.

Librarians have not strongly and consistently formulated their needs related to metadata in the past and this has led to a lack of action from other actors in the supply chain. Purchasing consortia (who negotiate licence agreements with content providers on their behalf) do not systematically make metadata provision an essential requirement; content providers do not always demonstrate their commitment to improving the quality of their metadata supply; system vendors do not live up to their promises, leaving librarians feeling powerless to unpick and fix the vast metadata flows they are largely separated from.

> 'librarians feeling powerless to unpick and fix the vast metadata flows'

We do not envisage that the metadata examples and landscape described in this paper will surprise many metadata and e-resource librarians, system vendors or content provider customer services. What we hope to achieve rather than just raising awareness is to reach an audience beyond the librarian who sees these issues every day. We hope to reach those

librarians who are affected by metadata problems but do not necessarily know why, and librarians active on the acquisition side of the library's operations (content and systems) who need to be aware of the issues affecting the products they will evaluate and choose. But we also know that we need to reach system vendors and content providers who do not get to see these issues as they happen for our users on the ground.

We believe that it is only by widening the conversation in this way that we can convince the metadata supply chain for academic publishing that these issues are problems for all stakeholders and that we must create an open forum for discussing, prioritizing and solving them together.

The situation we describe in this paper has practical impacts. Poor quality metadata has consequences for librarians and end-users which we will describe in two parts:

- Discoverability and accessibility

- E-resources management: understanding value.

Discoverability and accessibility

Publishers and, increasingly, library services platforms (LSP) are having to provide and maintain library metadata, but this situation has developed quickly and without an accompanying dialogue across the sector, which has led to an often perplexing experience for both librarians and end-users.

There is increasing recognition of these issues out there, as evidenced by services such as France's Bacon (Base de COnaissance Nationale) from ABES,[7] Japan's ERDB-JP (Electronic Resources Database-JAPAN) from University of Electro-Communications[8] and the UK's Knowledge Base+ (KB+)[9] from Jisc and discussions around the development of the Jisc National Bibliographic Knowledgebase,[10] but also by conference panels, listserv posts, and vendor and system user-group projects. It is easy, and often appropriate, for these discussions to become quite theoretical, but it can be difficult for those working with specific metadata in specific local environments to see how they can influence the bigger metadata workflow. Perhaps because librarians who manage e-resources come from varying library areas (systems, resource management, acquisitions and cataloguing, for example) there is a reluctance to declare and share issues experienced locally, so it may be useful to share some everyday examples to facilitate the development of common ground and priorities.

Creating and sharing case studies is much more time-consuming than we often expect and requires good documentation and follow-up, which can be frustrating and professionally risky – should librarians admit to their peers that they are not sure that their resources are actually all discoverable, for example? But airing our frustration has been useful, thus far, in developing engagement with fellow cataloguers and e-resource managers and has, surprisingly, led to invitations from content providers who are becoming anxious about the discoverability of their content to demonstrate our experience of their metadata 'on the ground'.

Publishers at UKSG noted the following, for example:

> 'Do we need to put our bibliodata elsewhere or is adding it all to ONIX'[11] enough?

> 'How do I know metadata has been updated in your [library] system?'

> 'We send data but get no feedback on whether metadata should be changed. If we send updated information there are no guarantees that this is used.'

Engaging with detailed workflows visually can be more compelling than reading dense description, especially for readers unfamiliar with the processes described, but we also feel that one of the more difficult aspects of cataloguing today is its inherently unstable nature which is leading us to use images as evidence more frequently. What any one person can

see in a system can depend upon profile permissions, institutional subscriptions and time of day (before or after indexing, for example), to name just some of the variables local to an institution, never mind those at a system vendor or content platform which also affect what any one person can see in their library management system (LMS) or catalogue. We are having increasingly to share screenshots of what we can see in order to trouble-shoot, so we felt it was appropriate to continue this practice in our article, both to engage as diverse an audience as possible and to demonstrate how important visual communication is becoming for resource discovery.

So, rather than continuing theoretically, here are a selection of Aberystwyth University Library's practical metadata issues, issues which remind us that while sharing problems may not exactly halve them, it can empower librarians to respond collectively to systemic challenges which exist at a scale they cannot effectively challenge as institutions or individuals.

'sharing problems … can empower librarians to respond collectively to systemic challenges'

What does 'poor' metadata look like, then?

While we, like many libraries and publishers, are particularly concerned with e-resource metadata, our first example involves print books because even librarians and publishers who claim to have no expertise in metadata are familiar with how to find some on a print monograph.

All institutions are aware of legacy metadata gaps and problems inherited from changing practices and personnel and migration loss, which leaves many resources without enough good metadata to facilitate discovery and identification or to match successfully in external systems like COPAC[12] or WorldCat.[13] Fewer institutions realize that they are sometimes still adding to this problem with poorly described *new* material, occasionally print and often electronic.

Contemporary print publications can slip into library catalogues through new acquisition processes (in this case ordered for a reading list) and onto their shelves with much worse metadata than libraries have traditionally demanded and expected because there is so little of it on the resource itself. Figure 1 shows two covers of the same work, published in 2015 and 2016 respectively. The resource on the left in Figure 1 from 2015 provides the sort of metadata we are accustomed to (as you can see also in Figure 2).
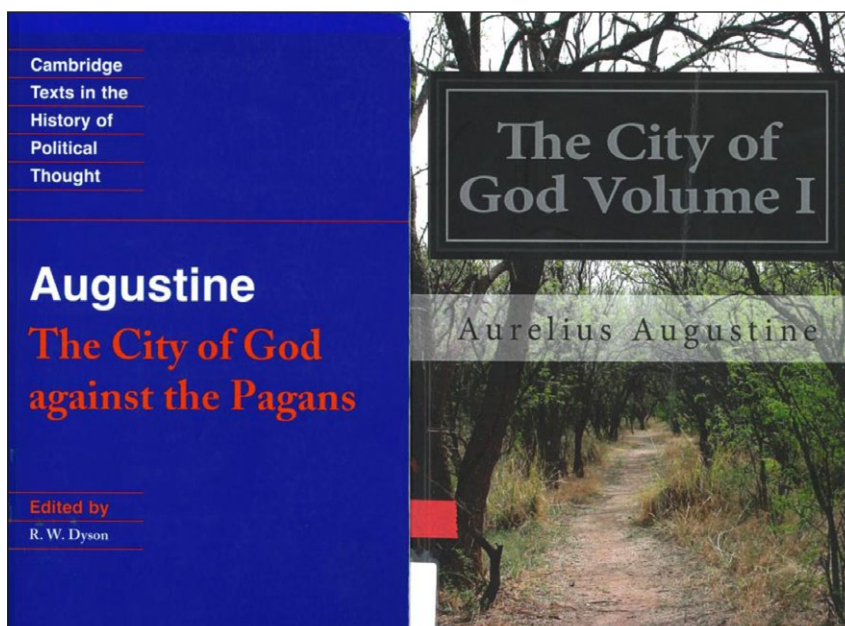

Figure 1. Two covers of the same work, published in 2015 (left) and 2016 (right)
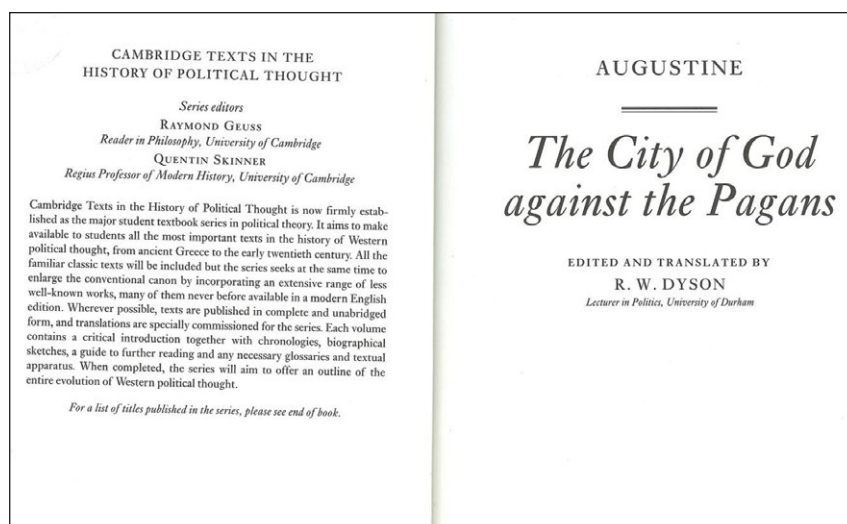
Figure 2. Series and series editor information facing the title-page with title, editor/translator name, institutional affiliation and publisher (2015 edition)

Continuing with this example from 2015, over the page, in addition to publisher, place and date of publication, there is edition information (first published in 1998, 5th printing) and copyright information, as well as ISBNs (Figure 3).
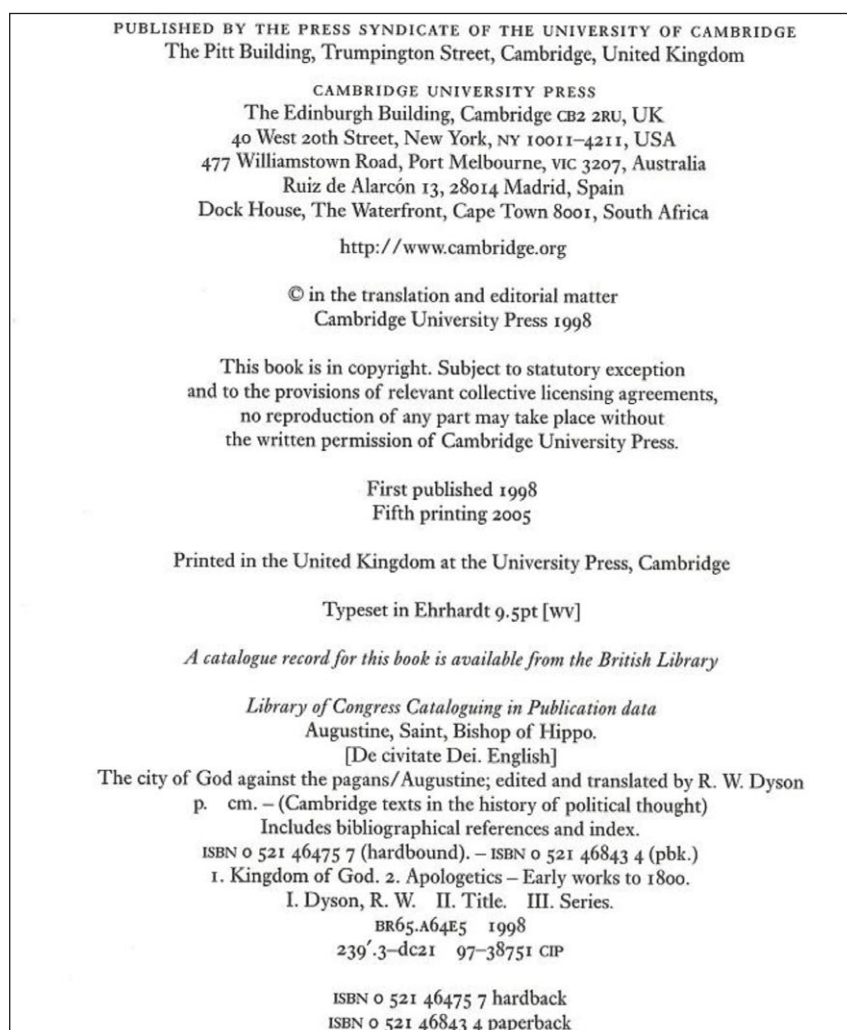

Figure 3. Edition and copyright information in addition to publisher, place and date of publication and ISBNs

6 The contents, which runs cross two pages (Figures 4 and 5), begins by telling the reader that there will be a preface as well as an introduction but there is also going to be a chronology of the author Augustine and a bibliographical note. The second page of the contents tells us that there will also be biographical notes and an index. The preface (Figure 5) outlines the critical history of the editor's approach so, in addition to metadata that describes the physical production of the volume, we have metadata that describes the intellectual production of the content. Finally, the introduction begins to provide historical context, in clear and accessible language (Figure 6).

Contents

vii

Figure 4. Beginning of the contents

Preface

This translation has been made from the critical edition of B. Dombart and A. Kalb, published in the *Corpus Christianorum* series (2 vols., Turnhout, 1955). Despite numerous – and inevitable – misprints, this is the best available edition. I have also from time to time consulted the *Corpus Scriptorum Ecclesiasticorum Latinorum* edition of E. Hoffmann (2 vols., Vienna, 1899–1900). In view of the great length of the text, and in keeping with the general principles of the series of which this translation is a part, footnotes have been kept to a minimum. For very extensive annotation, the reader is referred to the edition of J. E. C. Welldon (see Bibliographical Note). No abbreviations have been used which are not standard or self-explanatory.

Figure 5. The end of the contents, facing the preface

Introduction

St Augustine

Augustine was born at Thagaste in the Roman province of Africa on 13 November 354, to parents of senatorial rank. His mother, Monica, was a Christian; his father Patricius was not, although he was received into the Church shortly before his death. Augustine was brought up as a Christian catechumen; as was commonly the case in the fourth century, however, he was not baptised as a child. His childhood seems to have been full of unhappy experiences, especially in regard to his education;[1] but he writes of his mother with great affection, and is grieved by the memory of the pain which his youthful lapses caused her. In 370, he went to Carthage to study rhetoric. There, he lost touch with Christianity and acquired a mistress, who bore him a son called Adeodatus. As everyone knows, he regards himself as having lived a deplorable life as a young man, although he does not seem to have done much that we should now regard as very shocking. He read Cicero's dialogue called *Hortensius* (now lost), an exhortation to philosophy which fired his enthusiasm for learning; he was attracted successively to Manichaeism, Scepticism and Neoplatonism; he greatly admired the *Enneads* of Plotinus. Having taught for some years at Thagaste, Carthage and Rome, he accepted a position as municipal professor of rhetoric at Milan in

[1] One can only with sadness note his rhetorical question at Bk XXI,14: 'If anyone were offered the choice of suffering death or becoming a child again, who would not recoil from the second alternative and choose to die?' Cf. *Confess.*, 1,14.
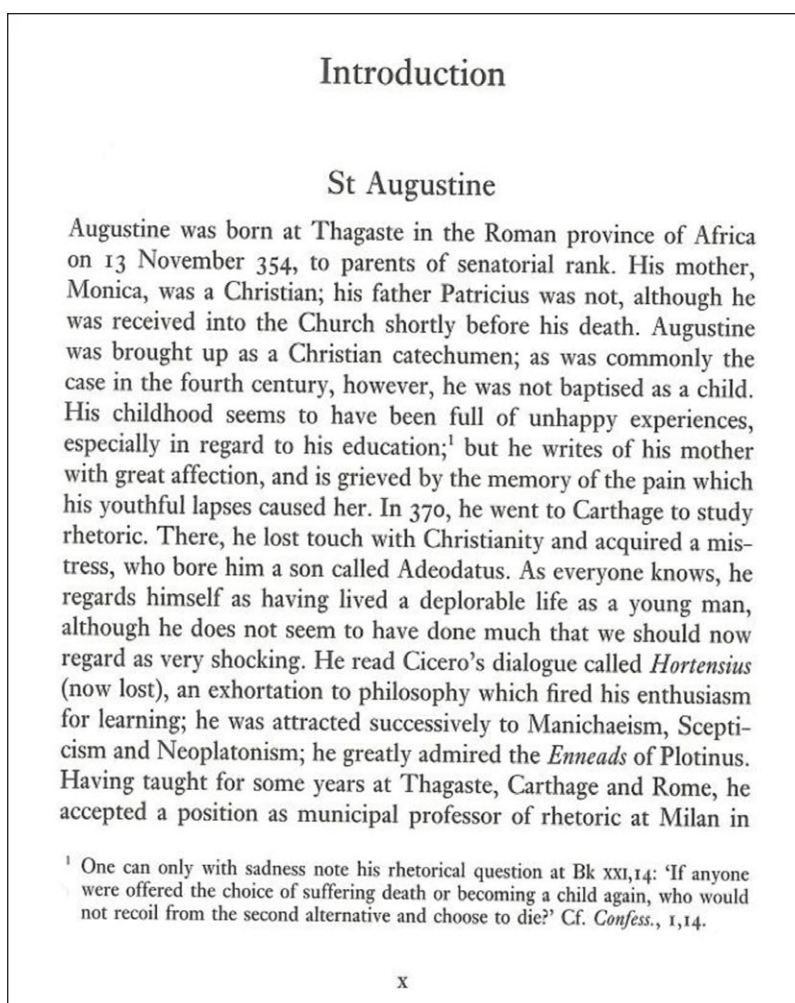
x

Figure 6. The beginning of the introduction

The resource on the right of Figure 1, published in 2016, looks quite different once you turn the cover (see Figure 7). It has a title page that tells us very little: no publisher, date or place of publication, and no explanation for Rev Marcus Dods' relationship with the work. (The remainder of the page is blank.)
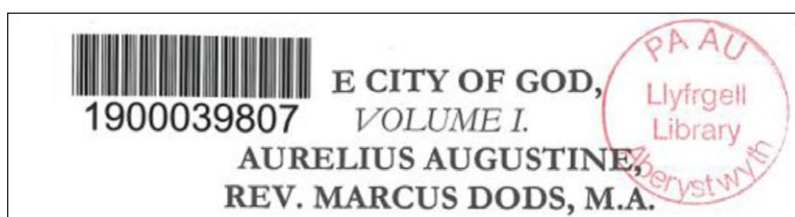

Figure 7. 2016 edition title page

The only other possible location for metadata on the resource is the 'Editor's Preface' (see Figure 8). We are not told who the editor is or when it was written. The text is small, the style old-fashioned and there is an assumption of knowledge, beginning: 'Rome, having been stormed and sacked by the Goths …'

With the more recent publication, end-users have nowhere to go if they are confused by the content (e.g. a contemporary introduction), no editor to cite in their essay or article, no publisher, edition or place of publication so that others can look up the information they have referenced, thus reducing the potential for scholarly communication. There is no sense of who edited this version and in what era, and the full meaning is affected by this omission because the work has been built, in this case, over centuries by different scholars, translators and editors, and their interpretations.

> **EDITOR'S PREFACE.**
>
> "Rome having been stormed and sacked by the Goths under Alaric their king,[1] the worshippers of false gods, or pagans, as we commonly call them, made an attempt to attribute this calamity to the Christian religion, and began to blaspheme the true God with even more than their wonted bitterness and acerbity. It was this which kindled my zeal for the house of God, and prompted me to undertake the defence of the city of God against the charges and misrepresentations of its assailants. This work was in my hands for several years, owing to the interruptions occasioned by many other affairs which had a prior claim on my attention, and which I could not defer. However, this great undertaking was at last completed in twenty-two books. Of these, the first five refute those who fancy that the polytheistic worship is necessary in order to secure worldly prosperity, and that all these overwhelming calamities have befallen us in consequence of its prohibition. In the following five books I address myself to those who admit that such calamities have at all times attended, and will at all times attend, the human race, and that they constantly recur in forms more or less disastrous, varying only in the scenes, occasions, and persons on whom they light, but, while admitting this, maintain that the worship of the gods is advantageous for the life to come. In these ten books, then, I refute these two opinions, which are as groundless as they are antagonistic to the Christian religion.

Figure 8. A section of the unnamed editor's preface

The fact that it has become much faster and easier to publish should not undermine the importance of what publishers provide that is unique and useful; context and accountability, not just content. And while we are sure most content providers would not produce a print resource with this little context and accountability, many librarians know that all too often this is how e-resources are presented to end-users.

Poor e-book metadata

It might seem inappropriate to compare a single print monograph with an electronic package as the latter are subscribed to because they offer access to vast numbers of resources, yet the same poor metadata can make the individual resource within the package totally inaccessible.

'The fact that it has become much faster and easier to publish should not undermine the importance of what publishers provide'

UKSG participants shared these concerns:

> 'Very small university with limited staff, we do not catalogue e-resources. We rely on [vendor] journal records and publication finder tool and find inaccuracies, inconsistencies, records change. We rely on website entries for databases (database A-Z).'

> 'Duplication of e-books records in [discovery layer] because the metadata is not consistent. Four records for the same book, for example!'

> 'LMS confused between same item on different e-resources.'

> 'We use System Vendor Knowledge Base … records are of variable quality. Cataloguers do e-book cataloguing and use [vendor] records. Varying practices leads to discrepancies in discovery/access/authentication practices in records.'

At Aberystwyth University, we were concerned that some system vendor knowledge base (SVKB) collections that we link to in the LSP to provide end-users with access to their e-resources were lacking in useful metadata, and the e-book shown in Figure 9 was the first example we followed up on. Figure 10 shows how this e-book looks in the catalogue for Aberystwyth University Library's end-users, and you can see that although it can be 'found' by searching the exact title, not a lot else would 'find' this resource. Even if you did come across it and wanted to be sure it was what you wanted, what could you use to 'identify' it?

Without being sure who wrote it, what edition it was, or even whether it was a book or a

LDR	----nas-a22----z--4500
008	170612uuuuuuuuuuuuuuu-|-o---u|---|eng-d
035	__ |a (CKB)4330000000079639
245	00 |a Essays on Some Unsettled Questions of Political Economy
260	__ |a |b Batoche Books
906	__ |a JOURNAL

Figure 9. Library management system view of e-book record with minimal metadata

Figure 10. End-user view of title search results list

journal (it is actually an e-book), would you 'select' it from other similar titles? Would you bother clicking through to 'obtain' it? As a librarian, if you wanted to do an e-book overlap test with another library's catalogue, would the record for this book be rich enough in metadata to match with another record for the same content? And as a publisher, would you be confident that your product was discoverable?

UKSG participants recognized this situation:

'Bad metadata and poorly catalogued e-books = no usage'

'Resources not found due to a lack of metadata'.

And yet, searching this exact title also brought back the resource shown in Figure 11, published in 1877, with author, edition information, physical description, contents and subject headings in the results list.

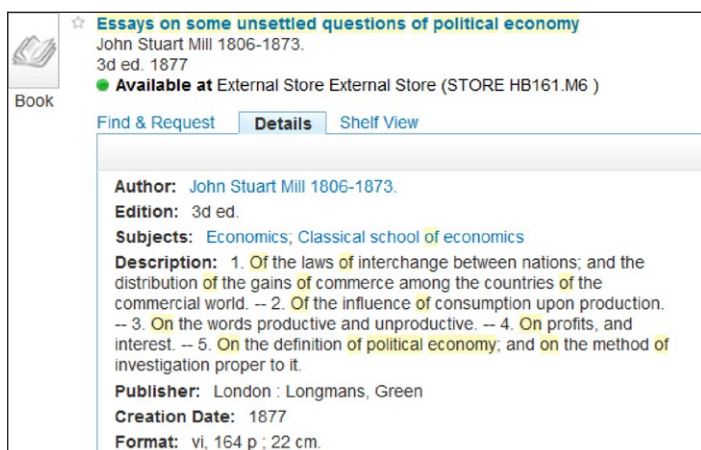The end-user is able, in this print example (Figure 11), to find the resource through a variety

Figure 11. A print source from 1877 in the title search for 'Essays on some unsettled questions of political economy' results list

of search terms and then to work out if this is the resource they need using author and title identifying information, chapter headings, place and date of publication or subject headings.

E-resources *themselves* may well have adequate or even good metadata but if that is not shared with libraries, or if it is shared and then overwritten by subsequent system processes that do not necessarily prioritize our expected metadata which can then be lost from the records, then the content becomes worthless.[14]

Poor metadata and reading lists

We found the above when testing our own concerns with e-book collections, but more immediately concerning are the issues that emerge from reading list management where we know end-users need specific resources at specific times.

Aristotle's *Poetics*, for example, was added by a lecturer to a reading list for Semester One 2017/8, so it had been available at some point. However, subsequent reading list link checks showed that it was later unavailable in our catalogue. We were able to find it on the publisher's platform where the metadata does include subject headings, a publisher and publication date, but no one is named as responsible for editing the resource, which, it turns out, would help end-users (and librarians and content providers) identify the work itself (as opposed to criticism about it) in our discovery layer.

But the book did not appear in the LSP or discovery service anyway, so our reading list link was broken. There are two similar titles in the same collection which are *about* Aristotle's *Poetics* but are not the text itself or the item linked to originally. (See Figure 12.)

At the time of writing this paper (in the summer of 2018 after Semester Two, the resource is still not available, although we have had a lot of discussion about the other two resources
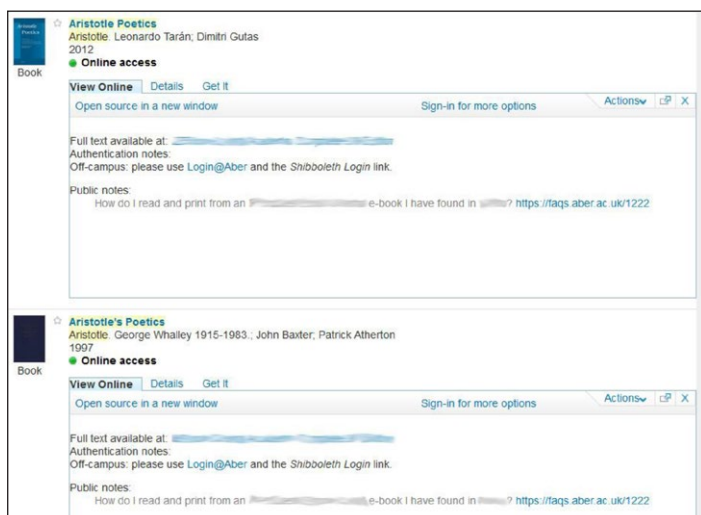

Figure 12. Searches for an original text may throw up results for books written *about* that text

which *are* available – both the system vendor and publisher keep saying we do have access because they find one of these two resources!

As with the print version, we have become aware of this item through our own checks but, unlike the print, we cannot improve the end-user experience in the discovery layer. And, worse than that, we do not know the scale of this problem because neither the publisher nor the system vendor can say with any certainty what we should have access to and in fact do have access to at any given point in a cycle of removals and updates. We have to rely on end-users telling librarians that something is not working to alert us that there is a problem.

So, this is an example of something that does not have great metadata for discovery in the first place, then does not appear at all in the discovery layer, and its lack of metadata makes solving the access problem then harder because it is difficult to 'identify' and 'select'.

This is an experience echoed by another UKSG participant:

> 'We have links in the discovery tool to [publisher platform] that go to a 'not found' page even though we have access to the content. Both [publisher platform] and our discovery provider say it is not their problem.'

The wrong metadata?

The *Journal of Volcanology and Seismology*, on the other hand, has a record with plenty of useful metadata for identification and is part of a large subscription that we link to through the SVKB. (See Figure 13.)

An academic member of staff got in touch because it is hard to 'find' and 'identify' the journal when the search result for '*Journal of Volcanology and Seismology*' and its ISSN comes back in our catalogue with an unexpectedly Russian title, although it links to the
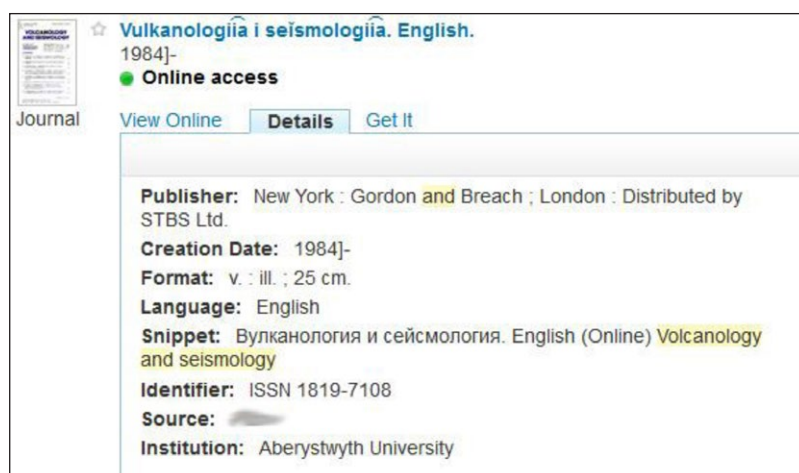

Figure 13. End-user view of e-journal record with Russian uniform title

English language content the end-user is expecting.

When we investigated the issue, we found that the SVKB record that we link to does indicate an English language version of a Russian text, starting in 1984 and published in the US and London, and its subject headings include a Soviet emphasis. There is another ISSN in the field for 'alternative format' information (the 776 field), so we wondered if this might indicate the existence of a different language version. (See Figures 14 and 15.)

We were unable to tell from a quick look at the resource whether or not the Russian emphasis was appropriate, so we looked the journal up in SUNCAT[15] where there were two aggregated records. (See Figures 16 and 17)

130	0_	a Vulkanologiia i seismologiia.	l English.
	0_	a Вулканология и сейсмология.	l English.
210	__	a J VOLCANOL SEISMOL+
210	__	a J. VOLCANOL. SEISMOLOG
210	__	a JOURNAL OF VOLCANOLOGY AND SEISMOLOGY
210	__	a J. VOLCANOL. SEISMOL
222	_0	a Volcanology and seismology
245	10	a Volcanology and seismology.
246	__	a Journal of volcanology and seismology
246	__	a Volcanology and seismology
260	__ |a New York : |b Gordon and Breach ; |a [London : |b Distributed by STBS Ltd., |c 1984]-

Figure 14. Library Management System view of e-journal record with Russian uniform title

650	_0	a Volcanology	z Soviet Union	v Periodicals.
650	_0	a Seismology	z Soviet Union	v Periodicals.
650	_0	a Volcanology	v Periodicals.
650	_0	a Seismology	v Periodicals.
776	__	x 0742-0463

Figure 15. Library Management System view of e-journal record with Soviet subject headings and alternative format field

The first, Figure 16, looked just like ours (with the same Russian information and dates of publication) but with a different ISSN.

```
022    $10742-0463$y0742-0463
030    $aVULSD9
035    $a(WaSeSS)ssj0055555
040    $aNSDP$beng$cNSDP$dDLC$dMH$dDLC$dNST$dInU$dNST$dNSDP$dMH$dOCoLC$dICU$dWaU$dOCoLC$dGBVCP$dO
CoLC$dOCLCO$dInU$dWaSeSS
041 1 $aeng$hrus
043    $ae-ur---
049 00$aSCID00020694814
050 00$aQE521$b.V842
082 04$a551.2/1/0947$219
130 0 $6880-01$aVulkanologiia i seismologiia.$lEnglish (Online)
222  0$aVolcanology and seismology
245 10$aVolcanology and seismology$h[electronic resource].
260    $aNew York :$bGordon and Breach ;$a[London :$bDistributed by STBS Ltd.,$c1984]-
300    $av. :$bill. ;$c25 cm.
310    $aBimonthly
336    $atext$2rdacontent
337    $acomputer$2rdamedia
338    $aonline resource$2rdacarrier
362 0 $aVol. 5, no. 1-
500    $aTitle from cover.
506    $aAccess restricted to authorised University of Cambridge users
515    $aVol. numbering adopted from Russian original.
516    $aText (electronic journal).
538    $aMode of access: World Wide Web.
580    $6880-02$aEnglish translation of: Vulkanologiia i seismologiia.
588    $aLatest issue consulted: V. 5, no. 6 (1988).
650  0$aVolcanology$zSoviet Union$vPeriodicals.
650  0$aSeismology$zSoviet Union$vPeriodicals.
650  0$aVolcanology$vPeriodicals.
650  0$aSeismology$vPeriodicals.
655  0$aElectronic journals.
765 1 $6880-03$tVulkanologiia i seismologiia$w(DLC)   80649621$w(OCoLC)5871649
776 08$6880-04$iOnline version:$tVulkanologiia i seismologiia. English (Online) Volcanology and
seismology
776 08$6880-05$iOnline version:$tVulkanologiia i seismologiia. English$w(OCoLC)609474267
776 08$6880-06$iOnline version:$tVulkanologiia i seismologiia. English$w(OCoLC)609474275
```

Figure 16. First SUNCAT record

```
022   $a1819-7108$10742-0463
035   $a(OCoLC)259528625
040   $aOH1$beng$cOH1$dCIT$dOCLCQ$dVRC$dDLC$dVPI$dOCLCQ$dDLC$dOCLCO$dUk
041 1 $aeng$hrus
042   $aukblsr
049 00$aSCID00021625957
050  4$aQE521.5$b.J675
082 04$a551.2105$223
210 1 $aJ. volcanol. seismol.
222  0$aJournal of volcanology and seismology
245 00$aJournal of volcanology and seismology$h[electronic resource].
260   $aDordrecht :$bSP MAIK/Interperiodica
264 31$a[New York] :$bSpringer US
310   $aBimonthly
336   $atext$2rdacontent
337   $acomputer$2rdamedia
338   $aonline resource$2rdacarrier
362 1 $aBegan with v. 1, no. 1 (Feb. 2007).
540   $aCurrent copyright fee: GBP34.00$c22\1$5Uk
546   $aEnglish translation of articles that appeared in Russian title: Vulkanologiya i seismolo
giya.
588   $aDescription based on: V. 2, issue 4 (Aug. 2008); title from journal information screen (
OhioLINK, viewed Sept. 30, 2008).
588   $aLatest issue consulted: volume 7, issue 4 (May 2013) (SpringerLink, viewed October 23, 2
013).
650  0$aVolcanism$vPeriodicals.
650  0$aVolcanoes$vPeriodicals.
650  0$aSeismology$vPeriodicals.
765 0 $tVulkanologiya i seismologiya
776 08$iAlso issued in print:$tJournal of volcanology and seismology
780 00$tBulletin of volcanological stations
```

Figure 17. Second SUNCAT record

The second, Figure 17, had our ISSN but described our resource as we would have expected, i.e. in English throughout, starting in 2007 and published only in the US.

The resource itself says it is available from 2007 and has two ISSNs – one for print and one for online – so at this point we thought that this was part of a known issue where print records have been used for online resources in the SVKB with only the ISSNs being changed. Digging a bit deeper on the journal web pages, we found that there is a Russian connection, however:

'Pleiades Publishing is an international company that has been working in the Russian market since 1991 … Springer Science+Business Media is a partner of Pleiades Publishing that distributes the journals…'[16]

But ultimately, without spending more time investigating, on balance we prefer the second SUNCAT record because it has the right ISSN but is also in English as the end-user expects, and states the publication start date as 2007.

We explained this to our LSP provider last year and were recently told that we are welcome to edit the record in the SVKB ourselves. We could do this, but this would be a global change affecting every single customer linking to the resource, and our changes could also be overwritten by other SVKB metadata updates. As a consequence, we have not felt able to take action to help end-users.

We have, however, started to draw up our own conditions for making changes to SVKB records, for when we might submit our own records for e-resources to the SVKB, and for when we will take the step to host resources and records outside the SVKB, precisely because examples like this make us feel powerless to assist end-users and we worry about the extent of such problems that we may be unaware of.

A participant at the UKSG breakout session described a similarly frustrating experience:

> 'Publisher acquired by another publisher (a big one!): We have been trying for eight months (a lot of e-mails back and forth) to get an accurate MARC records set for e-book collection purchased several years ago (and it is a small set 100–200 titles). Every time they e-mail the 'amended' set something is missing.'

At least with this journal title issue, we could unravel what may have happened. Sometimes you get a real oddity that just keeps evolving, without resolution.

Mystery metadata

We noticed *Career information and resources for Austria 2011* because it was one of a number of e-books deleted and added in collection updates by our SVKB, and then we could not find it in the catalogue or discovery service. We could not find it on the publisher platform either, because it turns out that the title in the bibliographic record is *Austria Career Guide*. So, we should at least have access to it in our collection as *Austria Career Guide* (but so far, do not).

Since the problem was identified, we have been given access via a different publisher collection (which really confused us!), again with the title *Austria Career Guide* although the resource itself is called *Career information and resources for Austria*. The record has no author or date or place of publication and wrongly claims it is in French and, although back when we presented at UKSG in April the record did say it was a book, it now says it is a journal. (See Figures 18 and 19.)

To access the e-book, the end-user has always been taken to a journal search page to search 'within this publication' to find a contents list for the book, which does make it look like a journal issue. (See Figure 20.)

LDR	----cas-a22----z--4500
008	180702uuuuuuuuuuuuu-|-o----u|----|fre-d
035	__ |a (CKB)3280000000002057
245	00 |a Austria Career Guide
260	__ |b Going Global
906	__ |a JOURNAL

Figure 18. The Library's view of the record as it appears in the system

Austria Career Guide
● **Online access**

Journal

View Online | **Details** | Get It

Title: Austria Career Guide
Publisher: Going Global
Language: French
Source: ▬▬
Institution: Aberystwyth University

Figure 19. End-user view of the title shown in Figure 18

It is unlikely that anyone would find this e-book but, even if they did, then working out what it was would not be easy and trying to access the e-book so confusing that end-users would be likely to give up altogether.
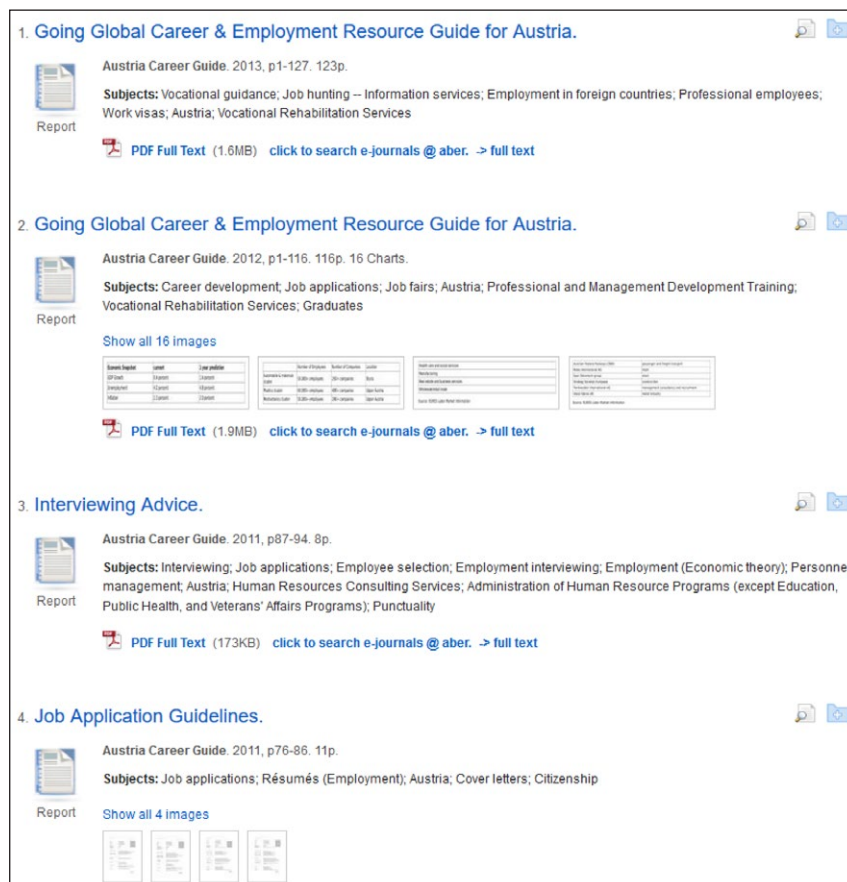

Figure 20. E-book chapters that looks like an article list

We were still trying to find out what had happened to our copy in the original collection. We were looking at other titles that had been removed and added, and we found that if you search for 'Career information and resources China' in our catalogue and discovery service, you get two results, one from 2004 and one from 2006. (See Figure 21.)

If you click on the 2004 China edition, it takes you to the 2011 edition of *Career Information and Resources for Austria*, with the title *Austria Career Guide* on the platform.

So, from an item being 'missing' we eventually work out that we have one title in two collections, both collections using the wrong title, one providing it through the wrong year and country, and the other as a journal in French with no date of publication.

Getting commercial companies – in this case two quite well-connected companies – to talk to each other and to us in a three-way conversation is proving incredibly difficult. We understand that, as a publisher/platform/system or aggregator, one resource amongst hundreds of thousands is a strange topic for discussion or focal point for problem solving. As customers, however, we represent end-users who experience specific resources not 'value for money packages', and we represent our collection management colleagues who need accurate metadata about our resources to take informed decisions – do we need to buy another e-copy or some print copies of this book for next week's lecture, for example? Librarians have to find a way to communicate this to content providers, and perhaps the concept of 'value' offers us a useful context for doing this.

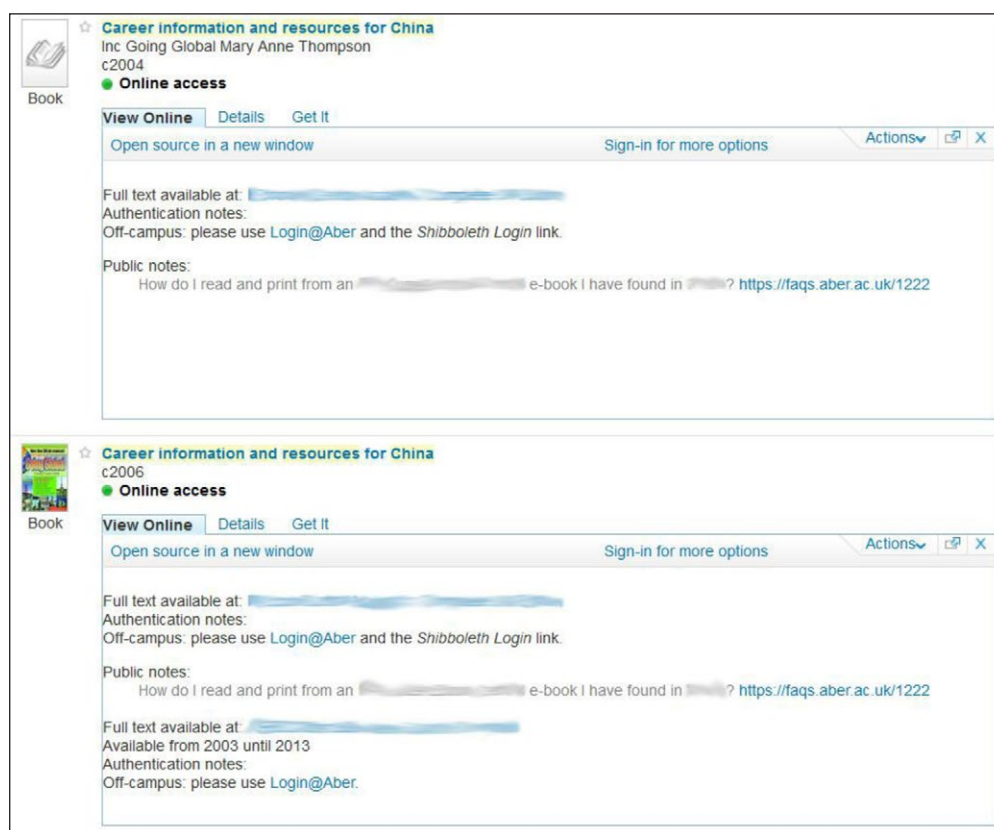'perhaps the concept of "value" offers us a useful context'

Figure 21. End-user view of title search results list

E-resources management: understanding value

One of the most demanding tasks for librarians is to realize the value for money of their collections. In a context of tight budgets and recurrent cuts, we need to be able to answer questions about the real extent of subscribed collections: what end-users have access to, the usage and the best acquisition methods to fulfil end-users' needs. The way that e-resources are sourced does not make life easy for librarians.

Content providers like to sell their products through bundles. Agreements negotiated by purchasing consortia include collections of journals and other e-resources and the volume of resources does suggest value for money when buying a collection as opposed to individual subscriptions.

Every renewal season content providers advertise their proposed bundles with phrases that imply that more content is being offered, with new titles, for example, or because the terms and conditions allow more access. And librarians do understand that they are gaining 'value for money'.

The content provider bundles can offer good products with terms and conditions that can be advantageous to end-users. However, we know that the bundles of items have a weak point, and this is metadata. The reality is that, in many cases, the bundles are not delivered with metadata that facilitate access and discoverability by the end-user and (importantly in this context) allow librarians to compare usage and value.

'when metadata describing the content of the bundle is not available, it is impossible to know what has changed'

We know that the particular resources contained within a bundle may vary, either at the end of a fixed-term agreement or even during the lifetime of an agreement, but when metadata describing the content of the bundle is not available, it is impossible to know what has changed. This can make comparisons of value over time impossible. Given the already tense situation regarding the cost of publications, this has the potential to lead to subscription cancellations.
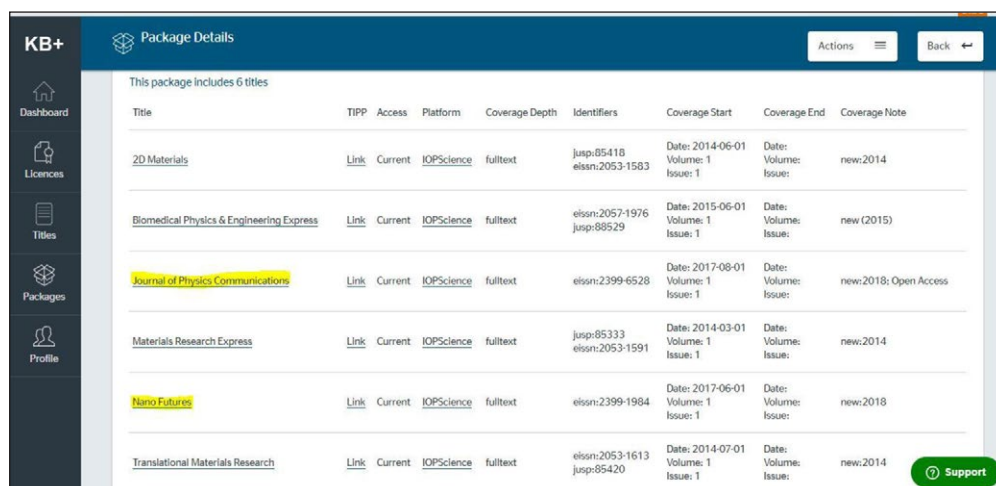
In 2013, to address the problems related to metadata provision, Jisc KB+ was established. The aim was to provide libraries with an opportunity to manage their electronic collections using a good set of metadata, and the KB+ team has excelled in the task as metadata provider. The metadata is not only available in KB+ but is also injected into the supply chain through open sharing and working with system vendors in their own knowledge bases. SVKBs consume KB+ lists and create specific targets. KB+ users then use these targets in the belief that they will see in their discovery services the same level of accuracy that they can find in KB+. But we have some examples that demonstrate that this is not always happening. These examples are real queries received by the Jisc KB+ team.

New titles not appearing in a knowledge base target

A librarian had a subscription to a bundle using a consortia agreement with the content provider for access to a specific collection. The agreement specified that new titles should be included.

As you can see in Figure 22, the KB+ team added these titles in the corresponding title list in November 2017 but, as of April 2018, the library's end-users who reported the issue were still unable to access or find these new titles.

Figure 22. An example of new titles added to a subscribed bundle by the KB+ team

This is not an isolated case; the Jisc KB+ team receives similar queries about other collections on a regular basis.

Titles deleted from a Knowledge Base target

The following example is similar, but it has been analysed differently.

Again, on this occasion, a librarian said,

> 'I'm checking through an xxxx update (from xxxx) for KB+Jisc Collections xxxx Full Collection xxxx, and the following titles have all been deleted.' [17]

The titles in question belong to a group of new titles that were promoted during the renewal agreement as a positive addition to the collection. In this case, we not only verified that the titles were part of the collection but decided to investigate if the same situation was affecting other subscribers to the same agreement. The results are shown in Table 1.

For this, we used a sample of ten institutions each with an active subscription to the same collection. For one institution, it was impossible to search their holdings so they do not appear on the results.

UKSG

Table 1. Titles available by system vendor

Journal	Institution 1 Vendor A	Institution 2 Vendor B	Institution 3 Vendor B	Institution 4 Vendor B	Institution 5 Vendor C	Institution 6 Vendor B	Institution 7 Vendor D	Institution 8 Vendor E	Institution 9 Vendor B
American Journal of Legal History	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Biological Journal of the Linnean Society	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Biology of Reproduction	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Botanical Journal of the Linnean Society	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Diseases of the Esophagus	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Endocrine Reviews	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Endocrinology	No	No	No	No	No	Yes	Yes	Yes	No
European Heart Journal – Cardiovascular Pharmacotherapy	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
European Heart Journal – Quality of Care and Clinical Outcomes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
International Affairs	Yes	No	No	No	No	No	Yes	Yes	Yes
Journal of Clinical Endocrinology and Metabolism	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Journal of Crustacean Biology	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
Journal of the European Economic Association	Yes	No	No	No	Yes	Yes	Yes	Yes	No
Neurosurgery	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes
Operative Neurosurgery	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes
Physical Therapy	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
SLEEP	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
The American Journal of Comparative Law	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
Zoological Journal of the Linnean Society	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes
Score	**95%**	**64%**	**36%**	**26%**	**78%**		**78%**	**100%**	**73%**
	band 5a	**band 5b**	**band 3**	**band 3**	**band 4**	**band 3**	**band 6**	**band 3**	**band 1**

The results show how many of the sample titles were discoverable in the libraries' discovery services. The results are based on simple searches and advance searches when the former produced no satisfactory results. The results were filtered by publication type 'journal'.

Each institution has been given a score based on the percentage of titles that were discoverable. Libraries use different system vendors that are connected to their own knowledge bases and link resolvers so, as expected, the results are different. Some of the institutions (2, 3, 4, 6 and 9) used the same vendor (Vendor B) but they also have different results depending on the targets they have activated.

The libraries belong to different-sized institutions and this is identified by the Jisc bands, which give an indication of library staff resources available.

Only one institution has a 100% score. The lowest score is 26% from Institution 4, where we found only five journals out of 19 were discoverable. For institution 6, it was not clear where their access was coming from, so we decided not to score them the same way.

We do not know how much the results are influenced by librarians checking and correcting problems manually, but Institution 1 did tell us that they spend time manually checking and correcting their link resolver targets.

'How many institutions are paying for items end-users will not be able to find …?'

This is just one example, but the problem is scalable. How many more titles are affected without our knowledge? How many institutions are paying for items end-users will not be able to find in their library catalogues/discovery services? How many disappointed end-users are there?

Both examples demonstrate that incorrect metadata has an impact on librarians and library end-users, which is experienced in a number of key functional areas of library service:

- inaccurate usage metrics

- inaccurate cost per use calculations

- impossibility of a complete understanding of the holdings

- potential for poor collection management decision making

- poor user experience

- poor user opinion of their library provision

- hidden costs for the libraries as librarians spend time correcting targets manually.

These examples show that the problem of inadequate metadata has its origin in poor provision from content providers but also from the lack of consistency in the absorption of metadata by system vendors (in poor updating and uploading of the data). Although these systems promise to reduce effort in managing discovery metadata, current experience shows a mixed picture. These are not isolated examples; Zhu has also shown the variable performance of system vendors when ingesting KBART phase 2 files.[18] The promise of pain-free resource discovery is simply not materializing.

'The promise of pain-free resource discovery is simply not materializing'

Conclusion

Let us go back to our original question, 'Is wrong metadata really bad for libraries and their end-users?'. The answer is,' Yes, it is'. There is enough evidence that inadequate metadata is affecting end-users and librarians. There is also enough evidence to demonstrate that the current situation is not just good metadata advocates making a fuss.

If progress has been made in the provision of metadata, with the adoption of basic standards or recommendations such as KBART, for example, it is not enough. More efforts to refine the provision should be made. A working metadata supply system can only be achieved if stakeholders step in and see themselves as owners of the problem. Clearly, each actor has somewhat different priorities, and makes a different balance of costs and benefits, but high-quality metadata benefits all stakeholders in the supply and use of library content. This will not happen unless we stop trying to solve individual problems and inject a component of openness and transparency into the metadata supply chain.

System vendors need to clarify their weak points and to plan changes in their development processes and resourcing. They need to communicate with librarians and content providers about their usage of the metadata they receive and, more importantly, the way metadata is prioritized, ingested, transformed and shared should not continue be treated as an 'industrial secret'. It should be openly available for consultation, reference and analysis by customers.

Content providers need to create and share good metadata but also to seek to protect it from degradation once released. They should not forget their original role of carriers of scholarly communication and let technology-related decisions dilute it.

Librarians should retake control of their needs regarding metadata and use all their interaction points with suppliers (content providers and system vendors) to reiterate the message. Libraries have the power as customers to influence vendors. Libraries can develop dialogue around the provision and processing of metadata, both pre- and post-procurement. The library community can use existing user groups and can move things forward by making sure that current customers' experiences reach system vendors and prospective customers too.

> 'the problem of bad metadata is too big and too widespread to continue with "small fixes"'

Discussions in silos – if they are happening between libraries, libraries and vendors, libraries and content providers, content providers and vendors – should all make way for an open conversation. We are not advocating for a naïve 'and we will all be friends forever' view. We are conscious that market implications and competition are behind all the stakeholders' actions. However, the problem of bad metadata is too big and too widespread to continue with 'small fixes' and marketing-related 'solutions'. All stakeholders will benefit from a forum where the exchanges of information will happen between content providers, libraries and system vendors. From the libraries' side, we could imagine that sector organizations such as SCONUL or Jisc could take the lead, showing a real commitment and willingness to stand for the solution.

The open question is: who will step up?

Abbreviations and Acronyms
A list of the abbreviations and acronyms used in this and other *Insight*s articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: **http://www.uksg.org/publications#aa**

Competing interests
The authors have declared no competing interests.

References and notes

1. Magaly Bascones was Knowledge Base+ (KB+) Service Manager at Jisc when the content of this article was presented at UKSG Conference, 2018.

2. Jeffrey Pomerantz acknowledges the stickiness of trying to define metadata when it itself becomes a container of data, albeit about other data. Like him, we'll settle for a working definition of 'metadata as a statement about a potentially informative object', *Metadata*, London, MIT, 2015. p. 26.

3. Greenberg J, Understanding metadata and metadata schemes, *Cataloguing & Classification Quarterly*, 2005, 40 (3–4), 20; DOI: **https://doi.org/10.1300/J104v40n03_02**

4. Hillmann D I, Metadata Quality: From Evaluation to Augmentation, *Cataloguing & Classification Quarterly,* 2008, 46 (1).

5. Culling J, *Link resolvers and the serials supply chain: Final Project Report for UKSG*, 21 May 2007: **https://www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf** (accessed 20 September 2018).

6. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional requirements for bibliographic records – final report*, 2009, p. 8:
    https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf (accessed 20 September 2018).

7. Bacon from ABES:
    https://bacon.abes.fr/ (accessed 24 September 2018).

8. ERDB-JP from University of Electro-Communications:
    https://erdb-jp.nii.ac.jp/en (accessed 24 September 2018)

9. Knowledge Base+:
    https://www.jisc.ac.uk/kb-plus (accessed 20 September 2018).

10. National bibliographic knowledgebase:
    https://www.jisc.ac.uk/rd/projects/national-bibliographic-knowledgebase (accessed 20 September 2018).

11. ONIX (publisher protocol):
    https://www.editeur.org/8/ONIX/ (accessed 12 September 2018).

12. COPAC:
    https://copac.jisc.ac.uk/ (accessed 20 September 2018).

13. WorldCat:
    https://www.worldcat.org/ (accessed 24 September 2018).

14. Another resource from the same publisher that has disappeared, Shakespeare's *Twelfth Night*, recently reappeared in our catalogue as *Twelfth Night: Arabian Nights (Chinese Edition)* published by Zhejiang Publishing United Group even though the record still resolves to the original Infomotion work (definitely not 'Arabian Nights' edition).

15. SUNCAT:
    https://suncat.ac.uk/search (accessed 20 September 2018).

16. Pleiades Publishing Group – About us:
    http://pleiades.online/en/publishers/about-publisher/ (accessed 20 September 2018).

17. After our presentation at the UKSG Conference in April, the content provider contacted us and let us know that they are working to address this problem.

18. Zhu J, Should publishers work with library discovery technologies and what can they do?, *Learned Publishing*, 2016, 30 (1), 71–80:
    https://onlinelibrary.wiley.com/doi/pdf/10.1002/leap.1079 (accessed 20 September 2018); DOI:
    https://doi.org/10.1002/leap.1079

Appendix

During the session we received the following feedback from the attendees; one more sign of the growing malaise that issues about metadata are creating and the increasing urgency for a change.

• [System vendor] accepting changes from libraries to their [system vendor knowledge base] without checking the requested changes.

• Our library unable to review weekly updates from [link resolver] sufficiently.

• Knowledge bases connected to discovery can be slow to update packages/collections.

• Information relating to changes to e-book collections for example can be slow coming from publishers.

• [System vendor's knowledge base] metadata.

• SHEDL deal metadata – large number of records – small team without time to check records.

• What is available in the journal collection at any given time.

• Some e-books have the date of digitalization or date of upload mixed with date of publication. Especially a problem on aggregators platforms that caused some old titles to be included that we don't want.

• From a publisher: Do we need to put our bibliodata elsewhere or adding all to ONIX is enough?

• From a publisher: How do we know if we are using the best possible keywords (+ how important are keywords – in addition to standard bibliodata to librarians and users)?

- KB+ and [discovery service] use different collection packages titles, difference in number of titles makes it hard to identify the correct collection or package to add to the discovery service [system vendor]. Others not up to date.

- Very small university with limited staff do not catalogue e-resources. Rely on [discovery service] journal records and publication finder tool, inaccuracies, inconsistencies, records change. Rely on website entries for databases (database A-Z).

- LMS has limited fields for licensing, supplier, admin information – inadequate as an e-resources management tool.

- [Library management system]'s user from Norway: Why does [system vendor] 'reduce' metadata from publishers when produce packages in knowledge base? How dare they do that?

- Supplementary material: indexed in discovery services but not indexed on publisher website so link resolver cannot accurately link to item on publisher site (conference abstracts, etc., published in supplements).

- I do have concerns that we often prioritize link resolving to the expected resource over the pursuit of accurate metadata – e.g. adding in ISBNS which match an object in the [link resolver] knowledge base to enable access where we are unsure of the veracity of the ISBN –local gains – but letting down community.

- Links in the discovery tool to [publisher platform] that go to a page not found even though we have access to the content. Both [publisher] and our discovery provider say it is not their problem.

- Chaotic update schedules.

- Discovery service surfaces some of our print records differently from their instance in our library catalogue. We catalogue them but they are 'ingested' and lose metadata. Please get in touch, I have more!

- Print serial holdings – accuracy.

- Bad metadata/cataloguing records on e-books – no usage.

- Difficulties in identifying proper titles on publisher platform and in link resolvers.

- Data about former publisher is nearly impossible to find.

- Cataloguing ditto makes link resolving difficult on article level.

- Duplication of e-books records in [discovery service] because the metadata is not consistent. Four records for the same book! for example.

- Resource not found due to a lack of metadata.

- Don't have time to check if we actually have access to the titles we subscribe to in the big deals.

- Legacy data from predecessor orgs – different format/structure and sometimes low quality.

- LMS confused between same item on different e-resources.

- As publisher: How do I know metadata has been updated in your system?

- From a publisher: system vendors – we send data but no feedback if metadata should be changed. If we send updated information – no guarantees this is used.

- OA titles in vendor sources – why not?

- Legal deposit libraries. E-legal deposit ongoing problems with poor article-level metadata from publishers causing access problems for users, in an already restricted access.

- Publisher acquired by another publisher (a big one!). Been trying for eight months (a lot of e-mails back and forth) to get accurate MARC records set for e-book collection purchased several years ago (and it is a small set 100–200 titles). Every time they e-mail the 'amended' set something is missing.

- Use [system vendor], serials use [system vendor's knowledge base] records variable quality. Cataloguers do e-book cataloguing – use OCLC records. Varying practices leading to discrepancies in discovery/access/authentication practices in records.

- Discovery layers (we use system vendor) different metadata … used between catalogue records and journal level (from knowledge base) + [discovery service] record – link resolver failures/[discovery service] not recognizing we have access to subscribed content.

- When reporting problems to [system vendor] I worry about the knowledge/experience/ competency of staff handling enquiry. Often I report a problem I say e.g. 'this is not right. I think it should be this. Can someone check + and if necessary/appropriate?' Not always convinced about competency of investigations.

- Delay in current year bundle record set appearing in [system vendor knowledge base] e.g. [publisher] SHEDL2018 set not yet present [system vendors knowledge base] and it is now April. How long do you wait? When do you chase? Who do you chase?

- Variation in whether different metadata/publisher list/title list etc. are using print or ISSN for a title.

- Want to see clearly what we have access to in full text, not what is available somewhere in the world … the years, the vols, issues, we subscribe to = viewonline.

- Using Syndetics Unbound – pulls in some strange information due to poor metadata in our records e.g. ISBN for a book in an e-journal catalogue record pulled in lots of unexpected data.

- Commercial/academic link resolvers supply of metadata very poor. Struggle to get title lists + links, let alone KBART.

- Aggregated subscription databases with content change.

- Discoverability of [content provider] resources in [system vendor's library management system and discovery service].

- [Publisher's] start dates! Delay of collection lists to [link resolver] + [library management system], etc.

Corresponding author
Magaly Bascones
Metadata and Discovery Manager, Digital Resources Division
Bloomsbury Publishing Plc, GB
E-mail: magaly.bascones@bloomsbury.com

ORCID ID: https://orcid.org/0000-0002-7092-2866

Co-author
Amy Staniforth

ORCID ID: https://orcid.org/0000-0001-7601-8133