How can bibliometric and altmetric suppliers improve? Messages from the end-user community

This article reports on a 2018 survey of bibliometric and altmetric practitioners – ‘Three things you want your metrics supplier to know’ – that was undertaken to better understand the practitioners’ usage of existing tools and services and to invite them to suggest ways in which they would like to see these improve. In total, 149 suggestions were made by 42 respondents, mainly UK librarians. Responses could be categorized into four main themes: A) Improve and share your data; B) Be more responsible; C) Improve your tools; D) Improve your indicators. The findings of the survey are discussed and sample comments shared. Based on these findings, and expanding on the four themes, the article makes a number of practical recommendations to metrics suppliers for ways in which their services could better serve the need of the community for robust and responsible bibliometric and altmetric evaluation.

Keywords
Bibliometrics; altmetrics; end-users; suppliers; tools; services

Introduction
The LIS-Bibliometrics forum was established in 2010 to support librarians as they started to grapple with an increasing range of bibliometric tools, indicators and enquiries from their academic communities. Since its inception, it has grown to be an active international forum of not just librarians but research administrators, planners, academics and suppliers. The forum commissions and undertakes small-scale research projects, such as that by Cox et al., runs regular practitioner-focused events and has established its own blog, The Bibliomagician. After a recent LIS-Bibliometrics event, some of the LIS-Bibliometric Committee members reflected that there seems to be too little engagement between the bibliometric practitioner community and an ever-increasing range of bibliometric and altmetric suppliers. The gap between their services and the community’s needs seemed to be widening. Were we the only ones that felt this way? How could we best start up a conversation with suppliers? Indeed, was there any consensus amongst end-users as to what their needs really were? The obvious first step was to ask them. So, using a methodology that had worked well for Clare Grace and Bernie Folan, the ‘Three things you want your metrics supplier to know’ survey was born.

Method
The survey invited end-users (e.g. librarians, research managers, academics and planners) to complete three free-text boxes outlining what messages they would like to convey to suppliers. A fourth open comments box was also provided alongside questions asking for basic demographic information about their country, job role and, optionally, their institution. A list of bibliometric and altmetric tools was presented and respondents invited to select those they used regularly. The survey was advertised through the LIS-Bibliometrics, ARMA-SIG-Metrics and RESMETIG discussion lists and promoted on Twitter. It was quickly
picked up on by some suppliers who were keen to get early access to the results, and they were helpful in further circulating the survey to their user communities. Responses were collected between 26 February and 26 March 2018.

The response data were coded, broken down into themes and then grouped to create some high-level messages. A second coder undertook an independent analysis of the responses and the two sets of outcomes were compared and synthesized to create this report.

Respondents

Forty-two individuals responded from eight different countries, the large majority (76%) from the UK (Table 1). Of the 42 respondents, 64% were librarians and 24% research managers or administrators (see Figure 1). The response rate would suggest that – just as with citation data itself – the user-base for bibliometric and altmetric tools is heavily skewed, with a small proportion of users making the most use of these services to the point that they had something really concrete to say about them. Although there were only 42 respondents, each provided up to four responses (three free-text plus one ‘other’ field). In total, 149 data points were collected, providing a rich seam of qualitative data for analysis.

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>32</td>
</tr>
<tr>
<td>USA</td>
<td>3</td>
</tr>
<tr>
<td>Ireland</td>
<td>2</td>
</tr>
<tr>
<td>Austria</td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. Countries of respondents

Figure 1. Job roles of respondents
Which tools were in regular use by respondents?

Figure 2 shows the bibliometric and altmetric tools most regularly used by respondents. It can be seen that Google Scholar and Scopus are the most frequently used tools (33 respondents), followed by Web of Science (29), Altmetric (27) and Scival (25). Fewer than 50% of respondents selected the remaining tools, although it was notable that almost one quarter of respondents (ten) declared themselves to be regular users of Dimensions, which had only been launched about a month before the survey was opened.

Findings and discussion

An overview of the most frequently occurring themes is given in Figure 3. These were grouped (where possible) to form four higher-level key messages, illustrated in Table 2. These key messages are explored in more detail below, illustrated with some typical comments. It should be noted that due to the nature of the research question around how suppliers could improve, the tone of many of the comments may appear negative. This should not be read as systemic unhappiness of all end-users with all aspects of suppliers’ products, simply the consequence of the issue under consideration. The full data set including all responses is available on the Loughborough University Data Repository.

Theme A: Improve and share your data

A1. We want greater coverage (preferably for free!), but if we can’t have that, please be honest about coverage limits

Respondents identified the limited disciplinary scope and coverage of different output types as a key barrier to satisfaction with supplier products. These issues were seen to undermine the credibility of the offer and to create unhelpful divisions within the academy – almost as if the exclusion of a discipline was seen as a kind of value judgment on it. As a result, there was some dissatisfaction with all the commercial products and services on offer as none of them had the breadth or depth to be universally useful. Users would clearly prefer some kind of ‘open Google Scholar’ (so long as someone else pays for the content and its indexing!). The long tail of additional disciplines and content types requested was very long indeed, and possibly uneconomic. Thus it would appear that incremental expansion is the best hope we have if we must stay with commercial
systems. Nonetheless, it would be helpful if suppliers could be more open and transparent about their selection criteria so at least end-users could understand their rationale.

Figure 3. Occurrence of themes amongst the data

<table>
<thead>
<tr>
<th>Message</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve and share your data</td>
<td>48</td>
</tr>
<tr>
<td>Be more responsible</td>
<td>41</td>
</tr>
<tr>
<td>Improve the functionality of your tools</td>
<td>29</td>
</tr>
<tr>
<td>Improve your indicators</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2. Four key messages to suppliers

Although expanded subject coverage was high on the wish-list, there might be unintended effects if this ever came to pass. Disciplines that are covered by citation benchmarking tools have seen an increased focus on publication in a small subset of highly cited journals that may subsequently increase in price. If coverage of Arts, Humanities & Social Sciences (AHSS) outputs in citation benchmarking tools increases, AHSS journals may be subject to the same fate, and scholars may find themselves under increasing pressure to publish in outlets that might not suit their preferred mode of scholarly communication. This might lead to inclusion of a kind they subsequently regret – especially if their outputs are included to a point that the coverage is passable, rather than laughable, but not to the point where it is sensible. Passable coverage may lead to worse evaluations than no evaluation at all. Any
large-scale expansion into new discipline areas would improve the apparent ‘citedness’ of those outputs and journals, perhaps giving a false sense of their growing impact. It would be important that suppliers are clear about these effects as outlined below.

Sample comments:

- ‘Coverage, especially in certain disciplines, means metrics tools are a very long way from being the one-stop shop they aspire to be.’
- ‘Improve coverage! The main barrier I have found to academics using metrics is the lack of coverage of any one database. We need a single, open source of metrics data.’
- ‘The Arts & Humanities are not covered well enough. You might be trying to get better coverage but it’s not happening fast enough.’
- ‘Highlight coverage warnings (A&H).’
- ‘Current bibliometric databases are missing a lot of the really important stuff for some departments, e.g. working papers.’
- ‘Perhaps a global, shared publisher portal with one access point to the stats.’
- ‘We don’t have the money to pay for any bibliometric or altmetric services.’

A2. We want better quality data (or at least be honest about its limitations)

Sound bibliometric analysis is utterly dependent on good data. As one respondent wrote, ‘accuracy is everything!’. However, respondents were clearly concerned that data quality is currently not satisfactory, especially considering the cost of citation benchmarking products. This is a serious issue because it undermines trust – and no one is less forgiving of errors than researchers themselves. Many institutions are reporting this information at a much more granular level than they ever have before, so this is becoming an increasingly significant concern. The clear message to suppliers here is ‘try harder’, which may be fair enough given the cost of their services. However, given that 100% accuracy is likely to be a) impossible, and b) extremely expensive (remember the long tail), the best advice here is probably for suppliers to be honest about data accuracy rates. If the author disambiguation rates are 95% correct, end-users could probably live with that as a rider to their analyses. If they are only 80% correct and the error rate is unquantified, then that is more problematic. Perhaps suppliers could publish data quality KPIs and numbers of correction requests just like the train services publish data on delays and cancellations?

Sample comments:

- ‘Data: more robust affiliation data, granularity, disambiguation, relationships.’
- ‘Data: better consistency in format and granularity of publication dates, including tidying up old publications metadata.’
- ‘Improve the quality of the indexing of authors and institutions’ profiles (e.g.: too many duplicates, spelling mistakes etc.).’
- ‘Improve author name disambiguation.’
- ‘Accuracy is everything! … I appreciate that there are millions of records but institutions are paying for it to be accurate, not to have to constantly report corrections.’
- ‘I believe that long-term consistency in approach to data collection is as important as the breadth of data collected. So start as you mean to go on.’
- ‘Data need to be as correct as possible, sometimes there are too many mistakes.’

A3. We live in a ‘mash-up’ culture – enable us to export, use and repurpose data

Universities are awash with management information and it was very clear from respondents that their role requires them to integrate bibliometric information with other data, not just to view publication indicators in isolation.
There may be some quick wins for suppliers here, for example ensuring that a standard set of identifiers is always available. This would be useful whether the data were a simple export of references from a citation database, the results of an online analysis, or the output from an added-value service like InCites or SciVal. Often, even within the same platform, it is not currently possible to join records because, for instance, one export format does not include an ISSN. The community also needs much more liberal system download limits and, more broadly, interoperability with a wider range of platforms, especially CRIS systems.

There seems to be a fundamental mismatch here between the perceptions of the suppliers (who seem to want to hardwire every possibility into their interface to 'make life easy for the user') and the reality (use cases are more complicated and sophisticated than they perhaps think; off-the-peg solutions often simply do not work).

Sample comments:

- 'Data we can export, transform and reuse in a transparent way is more important than pre-packaged proprietary visualisations and reports.'
- 'Make standard identifiers more generally available – DOIs for papers, ISSNs for journals. These are often available with normal downloads but not (e.g.) when exporting WoS/Scopus search analysis – just a list of titles, which makes reconciliation hard.'
- 'Allow the import and export/reporting of a unique identifier (e.g. a Pure UUID, or sequential unique range of values) in order to be able to better link input and output for further analysis outside of a tool/service.'
- 'Users often want to calculate their own metrics, compare information from different sources, or carry out their own subgroup analyses, not simply the “packaged” ones offered by the tools. A robust download functionality is essential for this. Download limits are often a major issue – increasing these would be really helpful. If there are concerns about abuse, perhaps a “basic download” function that stripped out the precise citation details but kept paper metadata and citation numbers would be a possibility.'
- 'The ORCID IDs of all our academics so we can properly track them.'
- 'Work with standards organisations to ensure interoperability of metrics e.g. CASRAI, euroCRIS, Snowball, ORCID.'

A4. Remember to whom the data belongs – a desire to reassert a sense of community ownership

Lying beneath all the calls for greater access to improved data was a strong sense from respondents that ownership of the citation record ought to belong to the scholarly community. Some respondents expressed unease that suppliers had better access to the community’s data than they do themselves. On these grounds it was felt that citation data should be opened up for the community to access, reuse and interpret. This is clearly the mission of the Initiative for Open Citations (I4OC) project, which describes itself as ‘a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data.’ It would be extremely helpful if all publishers provided cited references to Crossref on an open access basis for reuse. There was also a feeling that members of the community could be supporting each other to a greater extent by making available and sharing lists of researchers at departmental, school or faculty level to facilitate benchmarking.

Sample comments:

- 'Metrics belong to the scholarly community and should be freely available.'
- 'All publishers should open up their reference lists.'
- 'I believe we are in the dark in comparison to publishers in terms of gathering information about how our research is being used.'
- 'Institutions should be able to communicate/share openly, e.g. if they have created a group of researchers that represents one department it would be useful to share that with others to avoid replicating work and wasting time.'
Theme B: Be more responsible

B1. Suppliers have a duty of care to their end-users

Messages around the increasing importance of using metrics responsibly had evidently got through to respondents, but they were clear that this should be a shared responsibility with suppliers. ‘Metrics providers have a duty of care’, said one. This was particularly important when it came to indicators relating to individual researchers. As another respondent claimed, ‘It’s not their [own] fault that academics abuse metrics.’ In addition to many calls for suppliers to sign up to responsible use statements there were specific calls for particular indicators such as the h-index and Field-Weighted Citation Impact to be discontinued from individual researcher profile pages.

The individualized customer care and support offered by Elsevier was singled out for praise by a non-subscriber, although other respondents were more cynical about what they saw as ‘dishonest’ offers and felt that some suppliers were masquerading as a ‘beneficent uncle’ rather than the ‘profit-making company’ that they actually were.

Sample comments:
- ‘Metrics providers have a duty of care to the research community.’
- ‘There is nothing wrong with offering metrics solutions in academic contexts, but you should do this in a responsible manner.’
- ‘That signing DORA and/or publicly adhering to the principles of the Leiden Manifesto would be a positive step in the right direction.’
- ‘Remove h-index, FWCI and non-normalized metrics from individual researcher landing pages. These are not responsible metrics.’
- ‘As I said (and this will never happen) but I wish these providers were transparent that they make a lot of money and although I like what they can do, I feel like sometimes they are slightly disingenuous about why they are helpful.’

B2. Suppliers should provide better labelling for their products

It seems that end-users expect suppliers to enact their duty of care through better labelling and education. There was a clear message from the survey results that academics should not be held solely responsible for their own misuse of metrics (‘researchers don’t have time to appreciate the nuances’) and suppliers should therefore take greater responsibility. They should do so by making it very clear what their sources are, how the indicators are calculated and what their limitations are (e.g. sample sizes and confidence intervals). An analogy could be drawn here with food manufacturers. At a bare minimum, consumers want a list of ingredients (sources), but ideally they want a sense of how healthy those ingredients are, i.e. what percentage of our Recommended Daily Intake do they consist of (how sensible is it to consume these metrics, at what level of granularity, and what risk?); just as with food labelling, this could be colour-coded (and with error bars) if necessary. And if there are ingredients in there that could do serious harm, make it compellingly obvious – or even better, stop selling them at all (i.e. remove the h-index from researcher profiles). Just as producers of products that might be harmful are subject to higher rates of tax (sugar tax, anyone?) so perhaps suppliers should be tasked with investing a certain proportion of their income into education of end-users through the production of guides, training, promotion campaigns, etc. – but this is secondary and in addition to labelling the product correctly in the first place. Interestingly, the idea of using the Leiden Manifesto as a consumer label has also been explored by Wildgaard, Madsen and Gauffriau.9

Sample comments:
- ‘Easy-to-find and comprehensive list of data sources for that product.’
- ‘Please never let your products speak for them alone in form of simple counts, ratios, indices or rankings, without offering context information and interpretation.’
• ‘Stability/confidence intervals to contextualize indicators based on mean citations would be very welcome – it’d help us not to place too much emphasis on small differences.’
• ‘Rounding citation-based metrics to a sensible level would also help us not to place too much emphasis on insignificant differences.’
• ‘Add in error bars to any indicators so we can see how reliable they are.’
• ‘Clear description of methodologies and data used to develop metrics and allowing for independent (user) validation.’
• ‘More openness about underlying data (including description of weaknesses and for validation studies).’

... and on better education activities and use cases:

• ‘Real-world examples of how we need to use the data.’
• ‘Webinars and user group meetings are highly helpful.’
• ‘That good metrics require nuanced understanding, and researchers don’t have time to appreciate the nuances.’
• ‘Give guidance on how your metrics could be used in combination with peer review.’

Theme C: Improve your tools

C1. Find the sweet spot between innovation vs. the basics

Another cluster of comments complemented the more extreme mash-up sentiments above. There was some frustration that the current interfaces were not quite ‘right’ and this may be because suppliers do not really understand typical use cases well enough. Given the comments on scope and data quality, there may also be an issue here about the balance between getting the basics right, and constant innovation, often for features that are marginal to immediate user needs. Where products are designed for an international market, it is not always clear how the needs of those various markets are balanced against one another. It can be quite difficult when something that is seen as fundamental to one market (such as a date range that maps on to the current REF reporting period in the UK) is not forthcoming, whilst at the same time seemingly trivial ‘bells and whistles’ are introduced by suppliers, perhaps in response to an overseas market – or just because they can? The traceability of developments to the demands of particular user groups, and an understanding of their importance to that group, might alleviate some of the frustrations in this regard.

Sample comments:

• ‘I care about data quality: please invest in coverage and accuracy (even if it isn’t as glamorous as new developments).’
• ‘Underlying data quality is more important than flashy features.’
• ‘Mysterious ‘black box’ metrics and systems are not very useful to us – transparency is really important.’
• ‘I want intuitive and thoughtful UX.’
• ‘It’s nice when the interface changes but then it changes everything we do.’
• ‘More sophisticated visualizations, e.g. box plots, not just average values, for comparison.’

Theme D: Improve your indicators

D1. The ability to benchmark by small or niche fields would be highly valued

D2. Article-level subject indexing is needed

These two important issues are closely related. Subject fields are rather crudely defined in most bibliometric tools. An article is usually categorized by the journal in which it appears, which is ironically a fundamental no-no of most principles of responsible metrics. The call for subject indexing at article level was therefore an understandable one – although, hardly
without its complexities, as any librarian will tell you. Currently, a comparison between, say, Loughborough’s performance in the field of economics with that of King’s College, can only be done by looking at articles appearing in economics journals with either Loughborough University or King’s College as an affiliation. These articles may or may not have been written by individuals in the departments of economics at King’s or Loughborough, however. A further limitation is that filtering on economic titles will exclude economics-related papers in multidisciplinary journals.

To properly compare the two departments, you would need to plug in each individual working within those departments and/or their papers within an identical time frame, and run the analyses that way. Universities keep up-to-date lists of their own current staff, but not of peer institutions. One solution here may be for suppliers to facilitate the sharing of pre-defined groups between institutions, as suggested by one respondent. The challenge is further complicated when end-users do not want to simply compare one department with another, but a subdiscipline in one institution with a national or international benchmark. Being confident that you have identified all the correct individuals and/or papers is clearly extremely challenging without some form of article-level indexing.

Sample comments:

- ‘There is no good way to benchmark small departments, especially in niche areas.’
- ‘That each researcher and research project are different, homogenizing doesn’t work well.’
- ‘Research areas in departments differ a lot both across institutions.’
- ‘Improve filters so that system [sic] can filter at article level rather than journal level.’
- ‘Subject classification at the journal level isn’t clear enough. Subject keywords should be used to get better granularity.’

D3. Altmetrics are still nascent but better standards and integration would be welcome

There were very few direct comments about altmetrics despite 25 respondents stating that they regularly used Altmetric and 10 Plum Analytics. However, many of the generic comments may well have related to suppliers of altmetrics – especially those around transparency. The four comments specifically mentioning altmetrics called for a single standard means of collecting the data so that results from one tool can be compared with those from another, and a way of collating both bibliometric and altmetric data in one place.

Sample comments:

- ‘The variety of metrics is overwhelming; something that consolidates biblio and alt metrics would be great.’
- ‘Altmetrics is weak in my opinion and I’m not sure it is particularly useful except as a promotional tool. It needs to be developed into something which has a credible use and purpose, often when you drill down to look at the data or mention it is very weak or misleading.’
- ‘One altmetric standard – so that I can compare eggs with eggs. It must count web/social media sources, news and media mentions, and policy and grey literature mentions.’
- ‘We require the ability to view mentions from different sources simultaneously by being able to both select news sources and policy documents.’

Recommendations and conclusions

This survey provided a rich source of qualitative data around the needs and frustrations of end-users when engaging with the tools and services of bibliometric and altmetric suppliers. The key messages and recommendations are summarized below.
Theme A: Improve and share your data

A1. We want greater coverage (preferably for free!), but if we can’t have that, please be clear about coverage limits
• Suppliers should provide easily available, regularly updated lists of current coverage and signal more clearly any significant scope and coverage limitations.
• Suppliers should make it easier for customers to suggest new sources to plug gaps in disciplines and output types.
• Suppliers should make clearer statements on their plans for coverage expansion.

A2. We want better quality data (or at least be honest about its limitations)
• Suppliers should establish and report on KPIs around data quality improvement.

A3. We live in a ‘mash-up’ culture – enable us to export, use and repurpose data
• Suppliers should relax their system download limits.
• Suppliers should ensure that a standard and consistent range of identifiers is available for all data exports on their platforms to facilitate data integration and mash-ups.

A4. Remember to whom the data belongs – a desire to reassert a sense of community ownership
• Publishers should release cited references to Crossref under open licensing conditions.

Theme B: Be more responsible

B1. Suppliers have a duty of care to their end-users
• Suppliers should develop their own statements on the responsible use of metrics and adhere to them.
• Supplier should keep abreast of developments in the field of responsible metrics and update their tools accordingly.

B2. Suppliers should provide better labelling for their products … and better education activities and use cases
• Suppliers should provide clear, up-to-date guidance as to how indicators are calculated, with worked examples.
• Suppliers should provide clear warnings alerting end-users to the dangers of using certain metrics at fine levels of granularity, when the publication sets are too small to be statistically robust.
• Suppliers should provide confidence levels around their indicators, where appropriate.
• Suppliers should set aside a certain percentage of their revenue for educational activities to support responsible metrics.

Theme C: Improve your tools

C1. Find the sweet spot between innovation vs the basics
• Suppliers should involve end-users more actively in setting development priorities.
• Suppliers should provide a clear evidence trail for their development priorities.

Theme D: Improve your indicators

D1. The ability to benchmark by small or niche fields would be highly valued
• Suppliers should work to facilitate the sharing of benchmarking groups between members of the community.
D2. Article-level subject indexing is needed

- Suppliers should explore ways to develop more effective services (including enhanced benchmarking functionality) through output-level subject indexing.

D3. Altmetrics are still nascent but better standards and integration would be welcome

- Suppliers should integrate altmetric and bibliometric data to a greater extent.
- Suppliers should seek to standardize altmetric indicators and sources to better enable their interpretation.

We hope that these recommendations will serve to open up a dialogue with suppliers that moves us towards a better understanding of the art of the possible, and ultimately a more robust and responsible approach to bibliometric and altmetric evaluation.

Acknowledgements
The authors would like to thank members of the LIS-Bibliometrics Committee for their input into both the design of the survey and the writing of this report.

Abbreviations and Acronyms
A list of the abbreviations and acronyms used in this and other Insights 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