

# Rethinking Impact: Applying Altmetrics to Southern African Research

*Cameron Neylon, Michelle Willmers & Thomas King*

## Rethinking impact

A fundamental concern for institutions around the world is maximising the effectiveness and efficiency of their investment in delivering their mission; this is perhaps most acutely the case for institutions in the developing world that are forced to balance challenging social responsiveness agendas with the rigours of the increasingly competitive global higher education environment. This is often expressed as a concern to maximise the “impact” of funded research, where impact is intended to mean the effects of research beyond the research community. This might include influence on policy, improvements in health and living standards, cultural enrichment, or an improved environment. The emphasis on different forms of impact (and the framework by which that impact is assessed and rewarded) should depend on the goals and mission of the institution.

In a global political environment that places high value on transparency, accountability (Lao, Materu & Saint 2009) and demonstrable return-from-investment (Huisman & Currie 2004; Kruss 2012), tertiary education institutions are also under increasing pressure to provide evidence of the value of the services they provide. Traditionally, the provision of teaching

“

A key question is whether institutions in sub-Saharan Africa are in a position to meet demands for the improved measurement and dissemination of research outputs.

“

Thus, the importance of “grey literature” is massive, satisfying both the need for social engagement as well as scholars’ professional expectations.

services has been relatively straightforward to measure, with throughput rates, nationally/internationally accredited qualifications and provision of postgraduate qualifications being quantifiable measures of performance. In contrast, it has been less easy to measure and report upon the research mandate of universities.

Scholarly impact is nonetheless a crucial part of institutional evaluation, made all the more important by increasing competition in the academic sphere and the desire for institutions to appear relevant and provide desirable academic services. Quantitative measures of impact are useful and sought after as an “objective” measure by which institutions can plot their performance against regional and international competitors. The commercial publishing sector provides citation-based analysis (the ISI Impact Factor) which serves as the dominant metric for research evaluation, but the ISI Factor has come under increasing criticism for its methodology, equitability and ability to measure the complete range of scholarship. It is being challenged by the Altmetrics movement – a body of scholars and technologists that seeks to create tools allowing scholarship to be measured and tracked in novel ways.

The Scholarly Communication in Africa Programme (SCAP) set out to explore the current state of scholarly communication at four Southern African universities while probing the current level of alignment between the values of their academic communities, the mission statements of the institutions in which they work, and the reward and incentive frameworks that govern academics’ careers. This Altmetrics case study

comprises one component of SCAP’s investigation in this area. The objective was to speculate on alternative methodologies for a more Afrocentric approach to research evaluation that could align quality concerns, recruitment, recognition and reward systems in order to promote greater access to knowledge.

### Impact in the developing world

In a developing country context, such as in sub-Saharan Africa, there is a pressing need for developmentally-focused research. The historical deficit of higher education – due, in part, to the structural reforms dictated by the International Monetary Fund (IMF) and World Bank after the devastating effect of the global economic recession in the 1970s – has left sub-Saharan Africa with a shortage of skilled tertiary-educated professionals who can easily translate from basic research to applied socio-economic solutions. Thus, the importance of “grey literature” – policy briefs, working papers, media articles and other scholarship aimed at lay audiences – is massive, satisfying both the need for social engagement as well as scholars’ professional expectations. However, the global academic literature remains dominated by northern research and developed-world models that do not always take into account the specific socio-political environment prevalent in the sub-Saharan region. In such models, the journal article (and, to a lesser extent, the book and book chapter) reigns supreme; there is little space in the publishing industry for more socially targeted research outputs. In Africa, therefore, there is an imperative for locally relevant scholarship (and innovative scholarly formats) to be utilised to solve

local problems. African university managements understand the need for socially applicable research, as statements about the importance of locally relevant research and academic–community interaction appear in institutional mission statements across the region.

### **African institutional policy environment**

The under-funding of African tertiary education in the 1980s and 1990s had a deleterious effect on research production and policy. Institutions saw their share of global research fall, just as the international higher education sphere was ramping up its research production, sharpening its policy environment and developing research strategies to sustain the growth in research publication. Meanwhile, African universities saw their teaching loads explode, the loss of local academic labour to foreign universities due to brain drain (Te Velde 2005) and constant reduction in university budgets requiring them to do more with less.

The recovery period after this educational recession, fueled in part by a changing international funding sector led by reforms in World Bank and IMF policy that now recognised the value of tertiary education, saw the rapid creation of institutional policies that served the research agenda of universities. However, the need for these institutions to catch up to the developed world (which itself was slowly beginning to rethink core concepts in research strategy) has contributed to gaps, old-fashioned concepts and other flaws in the institutional policy environment in Africa. Aside from these structural issues, African academics themselves have had little time to acclimatise themselves to changing research policy landscapes, burdened as they are with a range of teaching and administrative loads that leave them with little time for research, and thus familiarisation with research policy. The traditional focus of African universities as centres of teaching – not prolific producers of research – has only contributed to this problem.

Current systems of rewarding and incentivising research in Southern Africa are based largely on the ISI Impact Factor. While this served as the first widespread quantitative research measure and retains some value as a means of measuring impact

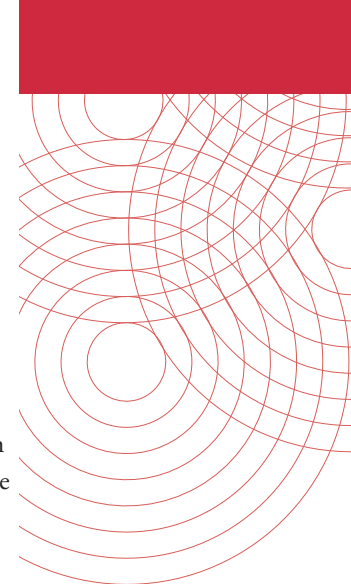
in the academic community, it is less suited to measuring other forms of impact, such as policy change, effect on clinical practice or economic benefit. New techniques and tools are being developed that can speak to these other forms of impact, but African universities persist (largely) in rewarding scholars based on their adherence to the ISI system.

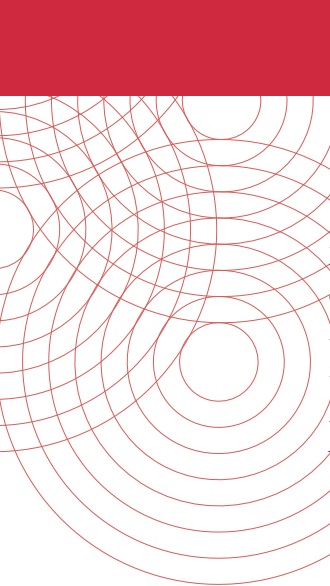
Nevertheless, tertiary education policy in Africa is not a homogeneous landscape. Within the continent there are institutions with solid policy frameworks for research, as well as those still developing institutional identities. New developments in international strategic research policy may allow African institutions an opportunity to stop playing catch-up and leapfrog to the forefront of innovative scholarship, if they are willing to embrace more open concepts of research impact.

### **Altmetrics: Realigning research practice with institutional values**

Research and its use and application are increasingly taking place online in a way that leaves traces that can be tracked and measured. In the wider business sphere many organisations exploit these traces to support strategic decision-making, especially the targeting of resources, for instance in advertising campaigns. This capacity to track the downstream use and influence of online resources, and especially the ability to obtain demographic information on the users of resources, has tremendous potential to aid institutions in tracking and testing the impact of their research activities and reporting to government, international funders and civil society.

These tools offer African institutions the ability to fulfil the developmental requirements included as part of institutional values and mission statements – an imperative made pressing by the many socio-economic problems that assail nations in sub-Saharan Africa. However, these tools rely on the substantive and critical reconceptualisation of the role of the university as publisher and disseminator of knowledge – a position that has not traditionally been foregrounded in academic practice (Cooper 2009). This expansive understanding of the





higher education sector's role in socio-economic development requires engagement by institutional managements in alternative publishing models, reward and incentives systems, and supportive policy, in order to fully realise the true impact of African research.

## Operationalising Altmetrics in Africa

Altmetrics contains the potential for a comprehensive reconceptualisation of what qualifies as impact, what should be rewarded in institutional reward and incentive structures, and how to track and promote engagement with civil society partnerships. Before Altmetrics techniques can be appreciated, though, “impact” has to be defined conceptually.

### Definitions of impact

The use of the word in the context of research arose in the UK where the government in the late 1990s and early 2000s was concerned with maximising the *economic impact* of research. This generated much resistance, with researchers and institutions pointing out that the value of research could arise from much more than simple economic considerations. As other forms of value creation were brought into the argument the narrow preoccupation with economic impacts was broadened and the term “wider impacts” came into use, often shortened to just “impact”.

Given that “impact” and the “impact agenda” are challenging political issues, it is sensible to question our use of these terms at all. The reason for adopting the term “impact” in this guide is that for better or worse, it is generally agreed to refer to the effect of research beyond the academy and captures the idea that there are a range of impacts of different types, which may be of different levels of importance to various stakeholders.

Various formal efforts have been made to define research impact, largely in the context of research assessment exercises. These various definitions all have limitations but their common feature is a focus on the extent to which a piece of research has


caused change. Examples include a modification to policy and the effect of that change, the extent to which a change in a clinical practice improves patient outcomes, or a measured improvement in water quality.

### Impact as use and re-use

Measurement of these kinds of impact is highly dependent on context and often very expensive, relying on a case-study approach. For the purpose of this guide we will adopt a simplified definition or, if you prefer, a simpler proxy of impact. Change in the world, whether within the research community or more widely, depends on research being *used*. Usage can cover many different things, from formal citation in the literature to a reference in a conversation, to incorporation in a textbook or a play, to policy briefs, environmental assessments or technology development.

We can choose to define impact as meaning use and re-use or we can treat use as a signal that provides evidence of “real” impact. A consistent feature of this guide will be that the data should never itself provide the answer to a question or the basis for a decision. Data can only ever support strategic decisions and should never be used as an excuse to avoid taking responsibility for a decision. With this in mind the question of whether usage *is* impact or is a signal of it becomes less important because in either case we will be using usage of information to support informed decision-making, conscious of limitations of our data and the frameworks in which we have chosen to collect it.

The advantage of a focus on usage is that with much of the consumption and discussion of research moving online, certain forms of usage can be tracked in ways that were not possible up to a few years ago. Another advantage of focusing on usage is that the measure on which much of our traditional tracking of research impact is based, formal citation in the scholarly literature, is also a measure of usage. The number of citations a research output receives is a measure of a certain kind of usage, by a certain kind of research user. By examining other kinds of usage we are simply measuring a greater diversity of kinds of use, and a greater diversity of kinds of users.



“ The advantage of a focus on usage is that with much of the consumption and discussion of research moving online, certain forms of usage can be tracked in ways that were not possible up to a few years ago.

## Measures of tracking impact

### *Citations*

Most quantitative analyses of the importance of research in the past have focused on citations. Some familiar measures, such as Impact Factors, Scimago Journal Rank or Eigenfactor are measures of journal performance rather than of articles. These are of no real value in measuring the importance of specific articles as the variance within articles in a given journal means that any average or aggregate measure is a very poor indicator of the individual articles.

However, the effort that has gone into these traditional measures now makes information on citations available at the individual article level. The number of citations a given article has received clearly tells us *something* about the article itself. In the spirit in which we will interrogate newer measures, we will now probe what citations can tell us.

### *What kind of usage do citations measure?*

The act of citing a scholarly work is a conscious signal from a researcher that a specific work has relevance to, or has influenced, the work they are describing. That is, citations are a signal that a scholar has used a specific article in their work. A citation does not necessarily signal agreement with the claims an article makes but it implies a significant engagement with the article. Citation is done purposefully and deliberately, meaning it is a measure that has some weight.

### *Who? What group of users does citation tell us about?*

Citation, as measured by citation services that are tracking the scholarly literature, is something

done by researchers, and specifically researchers contributing to the formal research literature.

Citations therefore measure usage by a very specific group for a fairly small range of purposes. It has a limited demographic reach and consequently a high focus on a specific community. This is both a benefit and a limitation.

With high-quality data it is also possible to provide more demographic detail on which researchers are citing a specific work. There are various kinds of demographic detail that might be of interest including geographical differences in interest, or awareness, career stage or prestige of the citing researcher, as well as the disciplinary areas from which citing researchers come.

### *Limitations*

Citations have a range of limitations, all of which can be recast as strengths. They are slow to accumulate, as they have to pass through the peer-review process to be registered. Second, they don't generally carry intent; it is seldom clear from the raw data the reasons why a specific paper is being cited. Is it to describe a method? To reference complementary evidence? Or is it to disagree or refute a claim? Finally, as noted above, they provide a limited view of usage and one that is almost entirely focused on re-use in research rather than application in the wider community.

By the same token the barriers to citation mean that they can provide a more measured judgement that accumulates over time, and their focus on the research community provides clear evidence of the importance of a piece of work in supporting further research. In some ways the lack of complex

information in the citation process has been crucial in making their collection tractable and the data useable.

### *Bookmarks*

As searching for and reading literature becomes an activity that is carried out in whole or in part online, tools for collecting and curating personal collections of literature, or just general web content, have become available. These tools make it easy, having found an article or web page of interest, to make a personal copy and also to build up a personal index of articles so they can be easily searched, managed or otherwise worked with. A side-effect of this bookmarking activity occurring online is that services can choose to provide information on the numbers of people that have bookmarked a specific paper.

There are a small number of services online that provide tools specifically targeted at researchers, with Mendeley and Citeulike being the most important for our purposes. There are other bookmarking and management tools, such as Zotero, Endote and Papers, but they do not provide any usage information to external users. Mendeley has the larger userbase and provides richer statistics. The data provided includes the number of users that have bookmarked a specific paper, groups that have collected a paper, as well as information on the demographics of users in some cases. The demographic information can include discipline, career stage and geography.

Bookmarks accumulate rapidly after the publication of a paper and because these services focus on research users they provide evidence of scholarly interest in a paper. This information is available earlier than citations in general and a number of studies have shown that bookmarks correlate reasonably well with the eventual number of citations.

In addition to scholarly bookmarking services there are also public bookmarking services such as Delicious and Diigo. They can provide a view onto wider interest in research articles. However, there is relatively little bookmarking of scholarly articles on these sites and the coverage of the literature

compared to Mendeley in particular is poor. This is in part due to users moving from these public bookmarking sites towards sharing via social media (below). Nonetheless, this provides our first clear case of a distinction between usage by researchers compared to the general public.

### *What kind of usage does bookmarking tell us about?*

Bookmarking of a paper is a purposeful act. It can be thought of as telling us more about the level of interest than a page view, but less than a citation, which demonstrates actual use. But bookmarking can also tell us about uses that are *different* to those captured by citations. In this sense bookmarks have the potential to capture uses that have traditionally been difficult. These might include papers that provide excellent background reading or introductory material, are position or policy papers that are important to researchers but often not highly cited, or are statements of community positions.

### *Who? Which users does bookmarking tell us about?*

Who a user of a bookmarking services is depends a great deal on the bookmarking service. There are services such as Zotero, Mendeley and Citeulike that are focused on the needs of researchers. Bookmarking on these services is generally done by people engaged either directly or peripherally with research. Each of these services in turn has a different profile of researcher users, with Mendeley and Citeulike being used more by science, technology, engineering and mathematics researchers, with a particular bias towards biosciences and computational biosciences. Zotero has traditionally had a stronger user base in the social sciences, humanities and information sciences. All of these services have a geographical user bias towards North America and Western Europe, a weakness they have in common with most online services.

As noted above there are general bookmarking services such as Delicious, Diigo, Pinterest and others. Not all of them provide useful data on the number of bookmarks for particular objects and in comparison with the researcher-focused services there is limited activity in most cases. The limited activity makes it challenging to determine whether signals from these services represent engagement with a wider public or their use by researchers.

Finally, bookmarking services can provide some demographic information – both in terms of career stage and geography. Mendeley provides information on the three most common countries in the bookmarks for a particular paper and some information on career stage. Citeulike, on which bookmarks are public, can enable correlation analysis – “readers who bookmarked this paper also bookmarked this one” – which can help to understand how papers are related.

### *Limitations*

The demographics discussed above also illustrate the limitations. Mendeley has good coverage of the biomedical literature with a high proportion of articles in PubMed having at least one bookmark. However, coverage is strongest in specific disciplines, making cross-discipline comparisons inadvisable.

The bias in the coverage of a bookmarking service is usually obvious from identifying the “top-scoring” papers. Bookmarking can only report on the activities of users signed up to a specific service, and this information is often not provided. For contexts outside North America and Europe this is an especially important limitation to consider. Finally, in common with citations, it is generally not possible to determine why a bookmark has been created.

### *Page views and downloads*

One of the major new sources of data available in the online world is the number of times articles are viewed. Page views and downloads can be defined in a number of different ways and can be via a range of different paths. Web views may be to abstracts or to full text (although for some publishers there is no distinction). We can distinguish between downloads of PDF files and views on the web; for articles that are also available in repositories like PubMedCentral, views at those sites can sometimes be provided.

Page views are an *immediate* measure of article usage as they can accrue as soon as the article is published (or if it is online prior to formal publication, even beforehand). Viewing a paper is an interaction involving less engagement than

citation or bookmarking but it can also capture interactions with a much wider range of users, especially if the paper is open access and therefore readable by the general public.

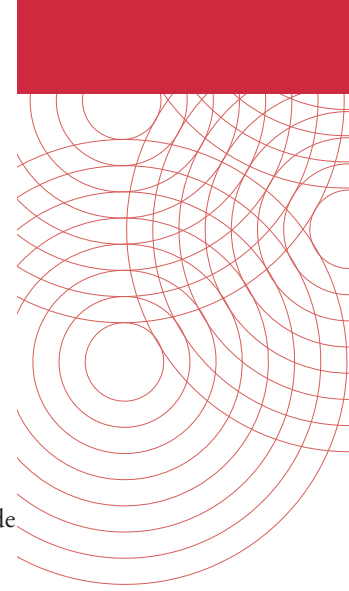
Some of the papers with the highest numbers of views are popular in the sense of having an amusing subject or title. Some are on issues of wide public interest and some are of more specialist interest. While it is easy to dismiss page views as being “merely about popularity”, this data remains valuable in two ways – first, in accumulating more rapidly than citations and, second, in providing evidence of types of use that do not necessarily lead to citations. For instance, a signature of important policy and position papers is that they are highly viewed, have significant but not outstanding numbers of citations, and have large numbers of scholarly bookmarks. Those that have captured the public imagination for other reasons tend to have high views, low citations and significant Facebook activity. It is the combination of measures that tells the story.

It is in principle possible to also draw demographic information from downloads but this is rarely provided externally. One example is the *Pan African Medical Journal*, which gives country-level information on those viewing an article.<sup>1</sup> This relatively underexplored area has significant potential for the future in providing detailed information on who is reading a given article, which may be valuable for determining, for example, whether research is reaching a specific target audience.

### *What kind of usage do page views and downloads tell us about?*

Page views and downloads provide information on the number of people who have arrived at a given article page or downloaded an article. It does not necessarily mean they have read it in detail or engaged deeply with the article. However, viewing or downloading is a prerequisite for most other forms of use.

<sup>1</sup> See, for example, [www.panafrican-med-journal.com/content/metrics.php?Src=Art&ManNum=2198&Vol=15&Issue=35](http://www.panafrican-med-journal.com/content/metrics.php?Src=Art&ManNum=2198&Vol=15&Issue=35)



*Who?*

Page views and downloads report on usage by those who have access to the articles. For publicly accessible articles this could be anyone, for subscription articles it is likely to be more focused on those with a professional research role.

*Limitations*

A key limitation of page-view data is that it may be calculated in a range of different ways and is not directly comparable across publishers. The data cannot easily distinguish between visitors who arrive and then immediately leave and those who stay on the page and engage more deeply. And if the paper appears in multiple places, such as an institutional repository or PubMedCentral, it can be non-trivial to re-aggregate all the page views.

*Social media (Twitter, Facebook)*

Social media are amongst the most valuable and least understood of the new services producing information about the use of research. When approaching social media tools like Twitter and Facebook it is most useful to step aside from the general hype and in some cases disdain for these tools and approach them from a new perspective.

In the context of understanding the usage of research, services like Twitter and Facebook are best thought of as ongoing coffee room or “water-cooler” conversations. While many of those conversations may be trivial, mentions of specific pieces of research are not uncommon. There is an increasing number of active researchers, policy-makers and technologists on these services and one of the things they discuss is research.

Consider for instance the following (real-world) example: a research paper<sup>2</sup> looking at the relationship between domestic violence and HIV status was published in *PLOS ONE*. By identifying Twitter traffic that included a link to the paper it was possible to show that the paper had come to attention of women’s crisis centres in Cape Town. Further, it was possible to understand how

knowledge of the paper had been distributed (via a researcher in Pretoria).

This example shows a number of features of social media as a tool. First, that it was possible to discover, amongst a large set of conversations, a specific discussion about a specific paper. Second, the nature of Twitter makes it possible to identify specific groups discussing the research and to confirm that these were potential targets of the research. Finally, it is in principle possible to reconstruct these discussions and to understand what path the research takes to potential users.

In future it will be possible to proactively identify target audiences, which might be geographical, disciplinary or demographic, and to ask whether they are being reached and how distribution might be modified to maximise that reach. This is potentially powerful, particularly for those involved in research with social or public health relevance.


The two major services that provide data are Twitter and Facebook – with Google+ and other longer-form writing sites and blogs also being useful in some contexts. Of the two, Twitter generally provides more useful data because the discussion itself and the identity of those involved are by default public. Those Twitter posts (tweets) that link to a specific article can generally be identified and connected to specific users. In turn the connections between users and the *other* things they say are often also available, making it possible to identify specific communities and to understand which ones are discussing a specific piece of work.

Facebook, despite frequent public scandals over privacy, has much less publicly available information and this makes it somewhat less useful. Specifically while Facebook provides general information on the number of people who have posted a link or “liked” a page that has a link to a paper, it does NOT provide information on the identity of those users or how they are connected into communities. It is possible to get some information about public pages and posts but this tends to be more limited than that available from Twitter.

As with bookmarking services, the demographics of the user base for Twitter and Facebook, and

<sup>2</sup> doi:10.1371/journal.pone.0024256





“ Social media are amongst the most valuable and least understood of the new services producing information about the use of research. When approaching social media tools it is useful to step aside from the general hype and disdain for these tools and approach them from a new perspective.

particularly the demographics of the subset of users that talk about research outputs, will bias any data available. Furthermore, the nature of the service may bias data. Facebook, being more private, can in some cases be a site for more frank discussion of a piece of research. Equally, the 140-character limit of tweets means that it is not a site that supports extended critiques. There is some evidence of strong geographical differences in the take-up of different services as well, although this often appears contradictory. In any case, social media are best looked at as a source of potential evidence for discussion and channels by which research might reach specific audiences. Absence of evidence should not be taken as proving a lack of discussion.

*What kind of usage can social media tell us about?*

As noted above, social media are best seen as a conversation taking place in a semi-public venue. Those discussing research are showing a level of interest in the research potentially greater than that involved in page views. Often users are simply passing on a link, essentially recommending that those in their community may find the article of interest. Generally speaking, mentions are associated with a positive recommendation but not always. An advantage of Twitter and to a lesser extent Facebook and Google+ is that it is often possible to navigate to the tweets themselves and to determine what exactly the level and nature of the interest is.

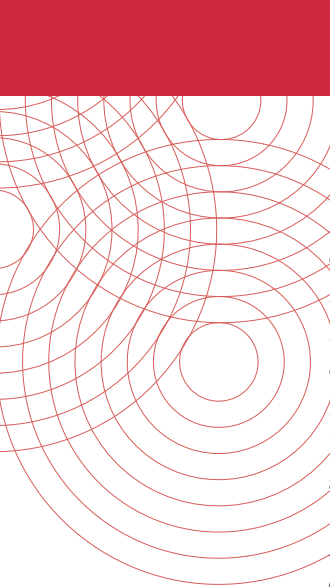
Social media conversations run the full gamut from highly technical and detailed to people finding the title of a paper amusing – sometimes in the same conversation. Thus numbers from social media data

sources should be treated with extreme caution and more detailed analysis is usually called for. There is a tendency for highly tweeted or Facebooked papers to also have significant bookmarking activity and to ultimately being highly cited, but the opposite is not necessarily true.

*Who? Which users do social media tell us about?*

The user bases for both Twitter and Facebook, the primary locations for online conversations and the source of most data, are global and public. There are strong geographical biases in the user base with South America and China in particular having a strong tendency to use other services. What is of interest here is the discussion of the demographics of those talking about research. A significant and rising proportion of researchers do use Twitter and Facebook in part for professional activities and this group is a major contributor to the volume of research discussion. Alongside professional researchers there are many journalists, policy-makers, public servants, public health professionals, civil society organisations and technology groups with an active presence on social media. Overall, most people coming to discussions of research on social media for the first time are surprised by the extent to which they involve groups of interested professionals.

If the professional discussion is a steady base level of conversation around a wide range of research, it is important to note that these discussions can easily be quantitatively swamped when a piece of research captures the interest of the wider public. When public message boards such as Reddit and Digg and popular websites such as Boing Boing



discuss a piece of research, the amount of social media activity around that research can easily increase by several orders of magnitude. The nature of this public interest obviously varies but it is important to distinguish between a piece of work “going viral” on the wider web versus a smaller-scale discussion by an interested community.

### *Limitations*

There is one important and serious limitation on the use of social media data. While it is generally straightforward to identify posts and comments that link to a specific paper, these discussions will often, perhaps generally, not contain such explicit links. The use of explicit links is likely to be biased towards those with a professional interest in research – therefore the data collection is likely to be biased against detecting wider interest from groups not directly engaged in research.

Alongside these issues, the same demographic issues seen for other data sources can play a strong role. These strong reinforcement effects – a paper that gets retweeted will reach more people, who in turn will likely retweet that in preference to other research – mean that naively relying on quantitative analysis of numbers of tweets or likes is not generally a useful approach.

## **Altmetrics tools and services**

Along with new ways in which impact may be measured have come new tools with which to collate these different forms of impact and present them in an accessible way. These tools have been designed to show a range of different impact measures based on individual research objects, thus allowing end users the ability to cross-compare different forms of impact for a single resource; extrapolated to an institution, this allows managers to see in which metrics their researchers are performing well, and in which they could use intervention.

### **Altmetric.com**

Altmetric.com is a subscription service that focuses on providing social media information for articles. It aggregates information from a wide range of social media as well as news sources. The focus


of Altmetric.com is on identifying those articles in a particular space with significant activity. The interface provides a range of search tools and more fine-grained control over which articles are selected. It is however possible to obtain data on a set of articles by providing DOIs, PMIDs or RePEC IDs.

Altmetric.com is helpful in probing social media activity and demographics. While ImpactStory does have demographic information in its database, it is not as easy to access and, as we will see, demographic information is very useful in identifying interesting stories. Another strength of Altmetric.com is the ability to create reports that are regularly run and can be sent to the user by email. This might be particularly useful at the institutional or funder level where a report can be set to run a PubMed search. Altmetric.com also provides a search function for a limited number of funders that are PubMedCentral partners.

Altmetric.com, similar to ImpactStory, suffers from limitations with Twitter data. Altmetric.com attempts to collect all mentions on Twitter of all published articles. This is stored by Altmetric.com locally. The resources and expenses associated with this is one reason why it is a subscription service. This means that Altmetric.com results, while being more comprehensive for more recent papers than ImpactStory data, are very limited in their historical reach for older papers. This however is common to many data sources that work in a similar manner. It is also the case that prior to 2010 there was much more limited Twitter activity around scholarly articles so the loss is less important than it might be.

### **ImpactStory**

ImpactStory is a free data aggregation service focused on helping individual researchers collect data around the usage of articles, datasets, presentations and other web content. It is also useful for institutional data gathering on a small scale. The service provides a web interface for creating collections of research objects as well as an API for more intense usage. For each object, the service will provide information on formal citations, bookmarking, social media activity and page view or download data where available.



“ Altmetric tools have been designed to show a range of different impact measures based on individual research objects, thus allowing end users the ability to cross-compare different forms of impact for a single resource.

ImpactStory is most effective for tracking the usage of articles where they have unique identifiers such as DOIs. This makes it less useful for books or popular media articles. It also relies on the user creating a collection. It therefore does not enable searching for relevant objects by name or institution.

ImpactStory reports can be stored and accessed subsequently for updating or tracking usage over time. The reports can be downloaded as CSV files for analysis in Excel or as JSON. The API provides results as JSON. The historical data is not accessible and there is not a regular reporting or alert function, so for those wishing to create a timeline, regular downloads will be required.

A key limitation of ImpactStory is its collection of Twitter data. Collecting comprehensive Twitter data for large datasets is potentially expensive and can require large resources. Twitter does not provide a mechanism for searching its archive for tweets more than a few weeks old. Because ImpactStory works on specific articles and seeks to gather data when a report is created or updated, it cannot search for historical Twitter data. Instead it uses another service (Topsy.com) that collates Twitter data. This means that ImpactStory is limited to the data that Topsy.com makes available.

### Other tools and data providers

#### *PLOS Article Level Metrics API*

For those willing to undertake a small amount of technical work another tool that can be of value is

the PLOS Article Level Metrics (ALM) API. This is the software that runs the PLOS ALM program but it can also be run as a stand-alone application. The software has been configured so that it is relatively easy to set up and it provides a mechanism for collecting data locally on a set of articles for which DOIs are available. Out of the box, the application provides data on citations (Crossref, PubMedCentral), Wikipedia usage, Mendeley bookmarks, Facebook activity, media mentions and usage information for a small number of publishers. For several of these a user will need to obtain their own “API key” for the source. This is generally a straightforward process.

The ALM API App provides a way of having data managed and collected locally, as well as a straightforward way of interacting with it. It does require some effort to maintain and work with, and the willingness to work on a command line, but it is well packaged and can be run by anyone willing to put in the effort to get it up and running. The reward is a system that is under local control and management and can be configured with exactly the set of articles desired.

The limitation of the ALM API is that it is focused specifically on articles and not other forms of research output, and that it is limited to a specific set of data sources. New data sources can be added if programming expertise is available, but for some data sources it may not be feasible to collect them locally. Twitter data in particular is very high-volume and requires significant resources to collect, making it impractical for a small local installation to manage.

### Research publishers providing usage information

A number of publishers and journals provide detailed usage information directly. Several publishers provide some information, generally including page views and usage data. This group includes the Frontiers Group, PLOS, the *Pan-African Medical Journal* and Nature Publishing Group. A range of other publishers provide information derived from either ImpactStory or Altmetric.com but do not provide usage data from their own internal systems. Some of these publishers also provide information privately to authors. Many also note or badge articles that perform well, noting that they are “highly viewed” or otherwise interesting.

### Case study: Altmetrics in Southern Africa

In order to experiment with Altmetrics in the African institutional context, the SCAP programme undertook a one-year content-tracking exercise in conjunction with the programme-participating institutions and an international Altmetrics content-tracking expert who operated as a consultant and principal data analyst. The three participating sites were the Southern African Labour and Development Research Unit (SALDRU)<sup>3</sup> at the University of Cape Town, a social sciences research unit with a focus on labour conditions and economics; the Faculty of Humanities and Social Sciences at the University of Namibia;<sup>4</sup> and the Faculty of Sciences at the University of Mauritius.<sup>5</sup> Due to the differences in context and discipline it was not considered appropriate to attempt to compare results across the sites. Instead, data was collected over a six-month period and examined in context to identify potential “impact narratives”, as well as to identify any interesting or unusual characteristics of the usage data.

<sup>3</sup> [www.saldru.uct.ac.za/home/](http://www.saldru.uct.ac.za/home/)

<sup>4</sup> [www.unam.na/faculties/humanities/humanities\\_index.html](http://www.unam.na/faculties/humanities/humanities_index.html)

<sup>5</sup> [www.uom.ac.mu/Faculties/FOS/index.html](http://www.uom.ac.mu/Faculties/FOS/index.html)

### Data collection

In order to conduct the content-tracking exercise, bibliographic data on outputs generated by academics in the SCAP pilot sites was required. Data collection took place over a six-month period, from May to October 2012, via institutional research coordinators who were tasked with sourcing lists of formally recognised institutional outputs for the period 2007–2012. It was assumed that these publication lists would be readily available within institutional research reports, but a number of challenges presented in trying to source the data.

Of the four original SCAP study sites, only two had publicly available outputs data for the five-year period. The third institution was able to supply data for a three-year period (they had only begun tracking research output in 2008) and the fourth institution was unable to supply the data despite repeated appeals on the part of senior management. The study was therefore confined to three participating sites. As the study sites within each institution were from different disciplinary fields, the research data collected varied from institution to institution. Traditional media (journal articles, book chapters and conference proceedings) provided the majority of research objects identified.

### Data processing and DOI ascription

As bibliographic information was received from each study site, a data entry and cleaning process was conducted. Bibliographic information was entered into an Excel spreadsheet, according to the CrossRef model of object ascription. OpenRefine<sup>6</sup> was then used in an attempt to normalise and clean the data. In particular, an attempt was made to normalise author names and the structure of bibliographic metadata. Where bibliographic metadata was limited or incomplete a further effort was made via manual web searching to complete or correct the information. Finally the available bibliographic metadata for journal articles was used to search against the Crossref database for DOIs for those articles.

<sup>6</sup> <http://openrefine.org/>

**TABLE 1: Bibliographic fields**

Publication type	First author	Second author(s)	Year	Title	Place of publication	Volume	Pages	URL	DOI
------------------	--------------	------------------	------	-------	----------------------	--------	-------	-----	-----

**TABLE 2: Number of object per digital handle**

Institution	Identifier	Number of objects	ImpactStory Collection
University of Cape Town	PubMed IDs	12	<a href="http://impactstory.org/collection/bd1ucm">http://impactstory.org/collection/bd1ucm</a>
	DOIs	48	<a href="http://impactstory.org/collection/bfuqab">http://impactstory.org/collection/bfuqab</a>
	URLs	39	<a href="http://impactstory.org/collection/y1ljzd">http://impactstory.org/collection/y1ljzd</a>
University of Namibia	DOIs and PMIDs	9	<a href="http://impactstory.org/collection/c2sljf">http://impactstory.org/collection/c2sljf</a>
	URLs	37	<a href="http://impactstory.org/collection/ur2kmz">http://impactstory.org/collection/ur2kmz</a>
University of Mauritius	DOIs	76	<a href="http://impactstory.org/collection/lesb5u">http://impactstory.org/collection/lesb5u</a>
	PMIDs	15	<a href="http://impactstory.org/collection/dxztal">http://impactstory.org/collection/dxztal</a>
	URLs	46	<a href="http://impactstory.org/collection/v3q9wl">http://impactstory.org/collection/v3q9wl</a>

Content was arranged by category, namely: journal article; book; book chapter; monograph; conference proceeding; poster presentation; doctoral dissertation; technical report; online document; review; newspaper article; public lecture; artwork; and other.

These categories were derived from institutional categorisation models and an effort was made to remain true to those structures. However, certain objects were placed into a category at the discretion of the PI team where their categorisation was unclear in the institutional reports themselves.

Once this preliminary process was complete, an attempt was made to find digital handles for each output, in the form of DOIs, RePEC IDs, PubMed IDs or, where this was not possible, URLs. Digital handles are essential as they attach to the article itself and provide a means to track its appearance through social media, indexing services and citation counts. For a portion of the outputs, no lasting handle could be found, especially with regards to artworks and technical reports. Success rates in acquiring handles was variable, from 78.8% to 22.3%. The variance rates in ease of institutional handle acquisition were positively correlated with the age of the institution and the journal article format (73% with handles, 48.7% DOIs).

### Collecting usage and impact data

For each site, a list of available identifiers (URLs, DOIs, PMIDs, RePEC IDs) was created. These were then submitted to the ImpactStory and Altmetric.com services to create reports. For ImpactStory the identifiers were divided to create separate reports due to the limitation of 100 objects in each collection.

These reports can also be reproduced by creating a new report using the relevant sets of identifiers from supplementary data for each site. Each collection was refreshed, run and a CSV file downloaded on 28 December 2012, 11 March 2013 and 29 March 2013.

For Altmetric.com a single report was created for each site containing all identifiers. These reports were run to collect all activity at any time. The result sets cannot be accessed as they are associated with a specific user account. The reports were set to provide a weekly email report and the full dataset was downloaded as an Excel file on 28 December 2012 and 29 March 2013. The email reports for the University of Mauritius are incorrect due to a bug in the reporting system. The reports for the University of Mauritius showed no activity. The activity that appears in the Excel file for the final report is a comment by the consultant on the challenges of troubleshooting a problem with Facebook data (see below) but does serve as a useful positive control.

## Analysis strategies

When faced with new kinds of data it is tempting just to rank items or to look for high values. While this is a way of potentially identifying interesting outputs, it is limited. Comparing this form of data across institutions or disciplines is also fraught with difficulties. It is often more useful to start from a specific context or question and to interrogate the data with respect to whether it sheds light on that question.

A question of interest in the context of a development agenda might be: “Is there evidence that this research is being used in a local context?” This allows the analyst to seek evidence of specific locales in the usage data, which might include citation information, social media or bookmarks. Social media, particularly Twitter, and Mendeley bookmarks can provide some information on the locality of users.

### *General comments on the data*

All datasets and collections containing journal articles showed some papers with significant numbers of citations and most had some with bookmarks. Social media activity was generally low with a small number of Twitter mentions. Generally, those items available only via URLs – reports and mainstream media pieces – received the most social medial activity with more items showing activity on Facebook rather than Twitter. Those items searched for by DOI or PMID showed virtually no Twitter or Facebook activity.

In terms of demographics, Twitter activity was generally too low to draw any significant conclusions. Twitter activity was dominated by tweets from users registering the USA or the UK as their location. Mendeley reader locations showed greater diversity with significant contributions from Africa (particularly South Africa), South America (Brazil) and Asia (India), but this varied from site to site.

### *University of Cape Town – SALDRU*

The South African Labour and Development Research Unit is an independent unit within the University of Cape Town that has focused on

studies of labour markets, human capital and inequality. The current focus is as a think-tank supporting government through high-quality social and economic research in South Africa, with a strong focus on public engagement.

The SALDRU bibliography was notable for having a significant number of RePEC identifiers in the set. These did not provide any useable data as the data services used were not configured to recognise them. These services have largely focused on science content and social sciences resources are poorly represented. An ability to determine usage of and activity around RePEC papers might significantly change these results.

Of the remaining outputs, journal articles provided the best data. These articles generated virtually no social media activity. What Twitter activity was evident appeared to be largely generated in the USA. A little over half of the articles (outputs with DOIs or PMIDs) had Mendeley readers. Readers were primarily based in North America and Europe but with significant representation from India, Brazil and South Africa.

This is a typical pattern for a research-intensive and conservative group of scholars undertaking relatively little wider engagement activity and having only a small engagement with social media. None of the Twitter activity captured appears to originate with SALDRU staff and most appears to result from either the journal itself or other allied groups.

The SALDRU data does not provide evidence of wider impact but suffers from two major weaknesses. First, the lack of identifiers means that the quality of the data that can be obtained is limited. Most of the data services available rely on identifiers to identify the full set of URLs that need to be tracked. Second, the nature of the work and the identifiers that are available are not well supported by the current data services. The data available is strongly skewed towards those outputs with DOIs and PubMed IDs, outputs that are perhaps not in the central domain of SALDRU work.

This illustrates an important point. While the overall picture shows little evidence of wider engagement with the SALDRU outputs, this does

not support a conclusion that there is no such wider engagement. Absence of evidence is not evidence of absence unless a good comparator organisation can be found. Even with a comparator it is likely that the current state of data availability is such that significant forms of wider interaction with SALDRU outputs would be missed. However, the paucity of social media activity does suggest there is an opportunity for SALDRU to make more use of social channels to promote its work. Given the increasing use of social media channels by government, including senior members of the South African government, there is an opportunity to achieve greater exposure for SALDRU work by using these channels.

#### *University of Mauritius – Faculty of Science*

The University of Mauritius (UoM) dataset, coming from a science faculty, was dominated by research articles and this provided a richer set of metadata to work from. Around half of the articles had DOIs available and this makes a significant difference to analysis. A significant number of the entries were conference papers that are more difficult to track or in journals that have a limited online presence.

The UoM dataset provided some of the starkest evidence of the weakness of data collection, with a number of URLs provided for journal articles at web properties that have not existed for some years. This journal had, in fact, moved to another site but neither the authors nor the university appeared to be aware of this based on the data provided. In some cases articles were provided which on inspection did not appear to include the claimed author.

For those articles published in African or other journals based in low- and middle-income countries, it was often difficult to find an online presence and the online trace of the article was therefore limited and difficult to follow. A clear divide emerged between those articles published in “international” journals by large North American or European publishers and those published elsewhere. It is not possible to disentangle the effects of the wider reach of these publishers, the greater ease of identifying and

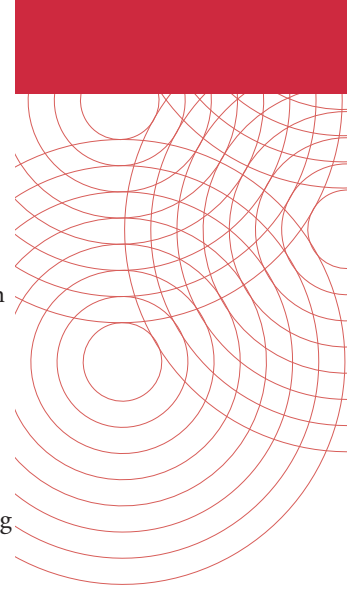
disambiguating activity around the articles they publish, publication “quality” and author selection bias. However, it is clear that both bookmarking and citation activity are concentrated around articles published by major publishers. For the UoM articles the number of Scopus citations was hand-transcribed to the dataset (this data is not provided in the downloadable data due to licensing restrictions).

There was essentially no general social media activity around the UoM articles. Three outputs have Delicious bookmarks, one article had three Facebook likes and one other had a single comment, like and bookmark. The apparent Twitter activity around one article is actually a comment from the consultant on a technical issue with resolving Facebook data for Taylor and Francis papers.

By contrast there is significant Mendeley bookmarking activity around those articles with proper identifiers. Out of 16 papers with PubMed IDs, 14 had at least one Mendeley bookmark. For those with DOIs, 31 out of 76 had some activity. This activity is greater than that for the University of Cape Town and the University of Namibia, reflecting in large part the disciplinary differences in the samples, as Mendeley users are concentrated in the sciences, and in the biomedical sciences in particular. Bookmarking shows a weak correlation with Scopus citations.

Amongst the most interesting data is the information on the country of origin of Mendeley readers. Due to the nature of the data it is difficult to provide a rigorous quantitative analysis. However, in the general Mendeley corpus it is unusual for articles to have the highest readership outside of North America or Europe. In the University of Mauritius dataset that pattern is quite common with India appearing commonly as the country with the greatest number of Mendeley bookmarks. In some cases a particular interest may be obvious: an article on the hygiene practices of rural food vendors in Mauritius<sup>7</sup> has readers in Malaysia and Nigeria; a study of anti-infective

<sup>7</sup> 10.1108/00346650410560361



properties of Mauritian plants<sup>8</sup> has readers in Indonesia. But there are not strong patterns: an article on the use of dietary and plant derived agents in chemoprevention<sup>9</sup> shows significant readership in India with US interest only equal to that in Malaysia.

It is difficult to draw any general conclusions but the data supports the idea that the diversity of interest in Mauritian research is greater than that for research originating in North America and Europe. If this is true it could be due to both the nature of the research itself, as well as the means by which it is communicated.

#### *University of Namibia – Faculty of Social Science*

The University of Namibia is a relatively new institution and has been tasked with the specific mission of developing policy briefs for the Namibian government. In this context it is interesting to consider whether evidence can be provided for local uptake of scholarly outputs. The outputs provided were a mix of journal articles, books, artworks and general media articles. As with the University of Cape Town data, it was journal articles that provided the most immediate additional information.

Three out of eight journal articles had citations. The articles with citations were those in “international” journals. It is not possible to identify whether these articles receive citations because they are in journals that are visible in the global North or whether citation activity is measured because the articles are published in journals for which citation activity is tracked. The sample is in any case too small to draw any statistical conclusions.

Many of the submitted journal articles were identifiable only as URLs. Generally this meant that only limited metrics could be obtained. There was however Facebook and Twitter activity around several of the URLs provided. As noted above our ability to dissect Facebook activity is limited and

<sup>8</sup> 10.1016/j.jep.2006.08.002

<sup>9</sup> 10.1016/j.tox.2009.10.010

this appears more prevalent than Twitter activity. Twitter activity occurs around both media articles<sup>10</sup> and journal articles.<sup>11</sup>

None of this activity appears to be directly connected to Namibia in a geographical sense. For the small number of articles with Mendeley bookmarks the top countries are the USA, UK, and The Netherlands. However, a Twitter mention of a governance article illustrates the potential for understanding how and why an article was accessed<sup>12</sup> illustrating that the article was available through a general search tool for articles on e-governance. Another article on “What is talked about when parents discuss sex with children” received at least one tweet by the Interagency Youth Working Group,<sup>13</sup> a US group focused on the reproductive health of 10–24-year-olds in developing countries. This was tweeted from an account with a wide range of followers, including many relevant groups and individuals from the African continent, providing direct evidence that the work was seen by relevant target groups.

As with SALDRU, a lack of evidence of wider engagement should not be taken as proving that such engagement is not taking place. As with SALDRU, work from the social sciences is not as well supported by the current data tools as work from science and technology fields. Nonetheless, there is a potential for driving greater engagement through social media and other channels and this may increase the exposure of this work to government within Namibia and the wider region.

## Findings

The quality of the data provided was generally poor, making it impossible in some cases to easily identify specific outputs online with confidence. Books and book chapters can be very difficult to trace and may exist only in print, and in some cases


<sup>10</sup> For example, [www.newera.com.na/article.php?articleid=41052&title=Repatriating+Namibian+sounds](http://www.newera.com.na/article.php?articleid=41052&title=Repatriating+Namibian+sounds)

<sup>11</sup> For example, [www.ajol.info/index.php/ajlais/article/view/63592](http://www.ajol.info/index.php/ajlais/article/view/63592)

<sup>12</sup> <https://twitter.com/cameronneylon/status/311814060373078016>

<sup>13</sup> [www.iywg.org](http://www.iywg.org)





“ It was clear that in each case the institution had resorted to requesting bibliographic information from individual academics, suggesting strongly that the institution has no central bibliographic store of its outputs.

journals had moved from one site to another. Some journal websites were also configured in such a way that articles were only accessible via a table of contents page for a specific issue.

These technical issues notwithstanding, it was clear that in each case the institution had resorted to requesting bibliographic information from individual academics, suggesting strongly that the institution has no central bibliographic store of its outputs. This is not unusual in a global context; high-quality and comprehensive recording of an institution's outputs is rare in UK and US institutions. Nonetheless, having a clear understanding of an institution's outputs is a prerequisite to measuring the performance and impact of those outputs.

*Finding 1: The participating institutions do not maintain good records of their own institutional outputs.*

In addition to the lack of centralised and comprehensive data collection it was also clear that researchers themselves retained very poor data collections. In Europe and North America most researchers will maintain a list of publications in the form of a document, usually providing sufficient reference information to find an output but rarely including detailed information on its online location. This is changing as online services like Mendeley, ORCID and ResearcherID provide the option for creating profiles online that link to the relevant outputs.

It was clear that most researchers did not have detailed bibliographic information or identifiers for

their research outputs. In some cases URLs were provided, but these were often incorrect and in some cases pointed to sites that had not existed for several years. Identifiers were rarely provided, with the exception of RePECs from SALDRU. In many cases, however, the bibliographic information was insufficient to easily identify the correct article or was incorrect.

*Finding 2: The participating researchers do not always retain high-quality information on the location and identification of their own research outputs.*

The cleaning and curation of the data to obtain as many DOIs, PubMed IDs and URLs as possible was an arduous and manual task. This probably involved over 75% of the time required to generate and analyse the usage data discussed here and represents an enormous wastage of resource. In addition to collating data on research outputs, institutions will be much better equipped to utilise available data sources if they can consistently collect persistent unique identifiers for those outputs where available, and relevant URLs where they are not.

*Finding 3: The bibliographic metadata available from institutions and researchers is poor and creates a very large workload in preparing a dataset. Systems that collect identifiers and online locations for research outputs will significantly reduce the workload involved in obtaining usage and impact data.*

Once a dataset was obtained and cleaned, using it to obtain information from a range of services was straightforward. Reports can be set

up and monitored with ease. These reports can provide insight into the use of research outputs, social media activity around outputs, and some information on the demographics of users. However, good data is highly skewed towards those outputs for which identifiers, particularly DOIs and PubMed IDs, are available.

*Finding 4: For outputs with DOIs and PubMed IDs it is straightforward to obtain data on use and performance. The available data is therefore skewed towards science, technology, engineering and mathematics (STEM) outputs.*

There was very limited evidence of social media activity around outputs. For each institution specific outputs had either Facebook or Twitter activity or both, but these represented a small proportion of the total. Twitter activity seemed to be largely related to North American or European accounts, often that of the journal, except in a small number of cases.

The lack of apparent local engagement on social media and the lack of evidence of engagement by the organisations themselves may represent an opportunity. Where the potential users of research outputs are engaged with social media, which may include local or international organisations, governments or other relevant local services, there is the potential for researchers and institutions to reach them more directly.

*Finding 5: There is very limited evidence of social media activity around the outputs from the three sites, and that activity is predominantly North American and European. Where potential users, especially local users, of research are active on social media, there may be opportunities for the three organisations to use social media to increase engagement with potential research users.*

Mendeley bookmarks provide some demographic information on the users of research from the three institutions. For SALDRU and the University of Namibia this shows a strong bias towards North American and European users. For UoM there was an intriguing tendency towards a greater diversity of readers. The limited sample and the nature of data means firm conclusions cannot be drawn

from this data, but the potential signal of greater geographic interest in science from UoM is worthy of deeper study. In all three cases the availability of data on Mendeley bookmarks is heavily biased towards both particular research domains and those outputs for which DOIs or PubMed IDs are available.

*Finding 6: From the available information on geographical distribution of users via Mendeley bookmarks, use of research for all three institutions is heavily biased towards North American and European users. For UoM there is a hint of greater geographic diversity but the limited data did not enable any firm conclusions.*

## Discussion

### The imperative for change

Internationally, tertiary institutions are struggling with alternative ways of measuring and tracking their own performance. The professionalisation of the higher education sector is bringing with it a new set of reporting and accountability standards that require more precise, more quantitative and more evidence-based proof of the value that universities provide to their students, countries and academia as a whole.

The power of the ISI journal ranking system has persisted, in part because until recently it has been the only way to quantitatively measure content, as well as the closed nature of academic discourse in which academics were primarily communicating with other academics. However, changes in higher education discourse have broken down some of the barriers between society and the academy, allowing academics to engage more directly with civil society, but also requiring that they adopt business practices that are endemic to that space.

A significant component of this new engagement is communication. Commerce and industry have eagerly seized the opportunities offered by social media in order to market themselves most effectively. Higher education will need to engage strategically with these new technologies and tools, not necessarily in the same manner as private individuals, but nevertheless with a mind to their



“

The lack of apparent local engagement on social media and the lack of evidence of engagement by the organisations themselves may represent an opportunity.

utility and the negative reflection that a lack of social media engagement will have on their operation.

### **The need for better data**

Throughout this study data quality was a significant issue. For institutions to make progress in exploiting the wider range of usage and engagement data currently available they will need to take a much more active role in collecting, collating and curating data on institutional outputs. While many North American and European institutions also have very poor data, they frequently have the resources to manage a collation process or the ability to buy in datasets from outside sources. The current lack of institutional infrastructure in the three institutions, while a challenge, also provides an opportunity to build up a workable infrastructure for collecting output data.

A key aspect of collecting outputs data is to obtain unique identifiers for these outputs. For working papers and preprints, the local identifiers should be collected. For published journal articles, DOIs and PubMed IDs should be collected where available and URLs required as a minimum. Institutions supporting or engaging with publishers that do not yet provide DOIs should consider the provision of DOIs as a priority. For books, ISBNs should be collected. While many of these identifiers are not currently supported by the data services that were used in this study, it is likely that they will be in future or that new services that utilise those identifiers will develop. Collecting as many identifiers as possible will create a data

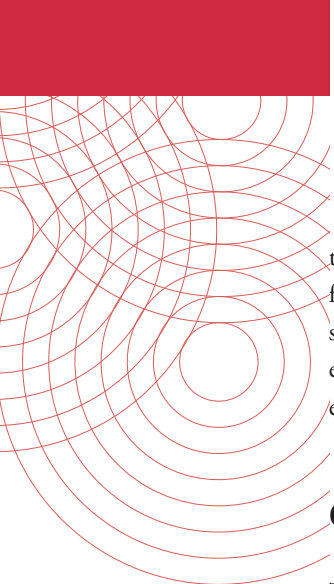
infrastructure ready to exploit future opportunities.

The collection of data and the preservation of a record of the organisation's output needs to be institutionalised. It is clear that relying on researchers to manage and provide this information is not reliable. This will particularly be the case where researchers move on. It is only through maintaining an organisational system for recording output data that this can be preserved reliably for the future. Such systems need not be complex but do require sufficient resource to be reliable.

### **A lack of evidence of engagement**

Throughout the study we saw very limited evidence of use or discussion of outputs. There were a small number of standout outputs with either high citations and academic bookmarking activity, or some social media activity, but rarely both. The availability of data was highly skewed towards biomedical journal articles which had DOIs or PubMed IDs and more frequently had academic bookmarks.

The limitations of the source data and the highly skewed data available means it would be unwise to draw a firm conclusion that there is only limited activity around these outputs. Nonetheless, the lack of a visible trace points towards opportunities for promoting work to potential users of research. Given the presence of community health and environmental groups, representatives of government and other institutions on social media channels in sub-Saharan Africa, there is at least a chance to reach key constituencies through



these channels. The apparent lack of activity from researchers themselves or their organisations suggests that it should be possible to increase engagement by taking a strategic approach and embracing these channels where appropriate.

## Conclusion

The SCAP Altmetrics investigation plays out against an international backdrop of wide-scale interrogation of the suitability of the ISI Impact Factor as the continuing sole measure of research impact. This situation has been amplified by a number of large research funders, particularly those operating in the development sphere, moving to disallow the Impact Factor data as evidence for research impact in grant fund application and reporting. As institutions around the world move to generate and engage with more sophisticated business intelligence for the purpose of better institutional governance and funder engagement, this case study raises important questions around the readiness of sub-Saharan institutions and research units to participate in this emerging area.

This case study serves to highlight an important disjunction in the relationship between values, mission and impact at the SCAP study sites. Attaining alignment across these areas while balancing the pressure to be locally relevant and working to attain international prestige will be one of the many challenges sub-Saharan universities face in the coming years. The Altmetrics tools

and methods discussed here provide a compelling means for African scholars and institutions to derive new forms of usage data on a wide range of research outputs to a broad spectrum of audience groups. In order to do so, further work is required in refining methodologies, ramping up institutional curation and research management efforts, engaging with research uptake strategies and growing additional research programmes to better understand the dynamics of implementing Altmetrics in Southern African universities.

## References

- Cooper D (2009) University-civil society (U-CS) research relationships: The importance of a “fourth helix” alongside the “triple helix” of university-industry-government (U-I-G) relations. *South African Review of Sociology* 40(2): 153–180
- Huisman J & Currie J (2004) Accountability in higher education: Bridge over troubled water? *Higher Education* 48: 529–551
- Kruss G (2012) Conceptions of higher education, development-oriented social engagement and innovation in the SADC context. In SARUA, *Perspectives on Regional Identity and the role of Higher Education in Southern Africa*. Johannesburg: SARUA
- Lao C, Materu P & Saint W (2009) *Legal Frameworks for Tertiary Education in Sub-Saharan Africa: The Quest for Institutional Responsiveness*. Washington DC: World Bank
- Te Velde DW (2005) *Globalisation and Education: What Do the Trade, Investment and Migration Literatures Tell Us?* London: Overseas Development Institute

