Are we comparing bananas and gorillas?

*Interpreting usage statistics for cost benefit and reporting*

Belinda Nicolson-Guest & Debby Macdonald
*Scholarly Information & Research, Information Services, Griffith University*
Background

- Introduction of Staircase Performance Management in 2011
- Scholarly Information Unit (Acquisitions/Vendor Services) provides 14 metrics for SIR (Scholarly Information & Research)
- Of the 14 metrics produced SIU provides 8:
  - 4 via in-house collected statistics
  - 3 via combined in-house & 3rd Party Usage Statistics
  - 1 via manual analysis of combined statistics
- These statistics are requested on a quarterly basis and confirmed annually
Today’s presentation

Three specific topics we will outline:

- In-house usage statistics
  ("stuff" we calculate)

- Usage statistics collected by a Third Party Service
  ("stuff" they calculate)

- Cost per download Statistics
  (Miracle happens here)
In-house usage statistics

- Time lag between end of reporting time and when publishers/vendors release statistics
- Non-Counter compliant statistics
- Statistics of non-traditional formats (e.g., video and images)
- Lack of statistics produced by vendor/publisher e.g., LUNA image collections

For example, both Adam Matthew Digital and Alexander Street Press (both excellent multi-media online resources) only provide DB1 Counter statistics (e.g., searches and sessions by month and database) which does not offer more granular statistics for image or video usage
Ebooks without usage statistics

- Despite the upward trend in ebooks purchasing, some providers still do not provide usage statistics. For example:
  - EOLSS – Encyclopedia of Life Support Systems (Unesco)
  - Encyclopedia of Pharmaceutical Technology (Informa Healthcare)

Extract from “QULOC IRA e-resources statistics trends” presented QULOC University Librarians Forum November 2012
Third Party Statistics

- Multiple means of access may result in some resources having higher relevance
- What constitutes a journal
- Changes in publishers and/or platforms
- Changes from p to e
- Design of platform
- Peaks and Troughs
- Types of articles
- Other Anomalies (eg. normalising title variations)
Both of these can be counted as a JR1 (full-text article request by journal title, platform and month)

In brief article from 20th November 2012

7 year longitudinal study by K.Ritchie
Duplication (normalisation of titles, etc.)

Extracted data from JSTOR

Extracted graphic from Serials Solutions Webinar
Cost per Download Statistics

- Rolling front files
- Full-text aggregator usage statistics
- Backfile statistics (to deduct or not to deduct)
- Exchange rate fluctuations
- Library systems vs. finance systems (who’s actually paying!)
Exchange Rate Fluctuations

US dollar volatility

UK Pound volatility

USD 2012
USD 2011

UK 2012
UK 2011
Conclusions

- One year of statistics is a snapshot in time
- Trend data creates a complete picture
- Usage statistics are part of a larger context for decisions
- Strategic intent of the institution is the larger context:
  - Research goals
  - Learning & Teaching goals
  - Student Experience

“Trends reveal the most about services and performance”

“Metrics are a means to an end, or, rather a means to a trend”

Usage Statistics: Bananas or Gorillas
Top 5 Usage Statistics Issues for Griffith

1. Lack of more granular statistics
2. Lack of consistency with the ebook usage statistics provided
3. The implementation of discovery layers has skewed the figures
4. The delay in third party providers collecting/making statistics available
5. Non-Counter Compliant statistics

Follow this link to participate in a short survey:
http://www.surveymonkey.com/s/FR95NQR

Deadline for survey is 31st March, 2013
Reference List


Thank you

★ Special Thanks to:

- Debby Macdonald (co-writer and chief editor)
- Anne Fleet (statistics guru)
- Tara Anderson (handout design)
- Christine Gordon from James Cook University (email advice)

★ Email contacts:

- Belinda Nicolson-Guest (b.nicolson-guest@griffith.edu.au)
- Debby Macdonald (d.macdonald@griffith.edu.au)
Any Questions?

Follow this link to participate in a short survey: http://www.surveymonkey.com/s/FR95NQR
(Slide 1)

Are we comparing bananas and gorillas? Interpreting usage statistics for cost benefit and reporting.

Presentation by Belinda Nicolson-Guest and Debby Macdonald

Introduction (Slide 2)

In 2011, Information Services at Griffith University introduced the Staircase performance management framework in order to align with the reporting undertaken by the Schools within the university. The Scholarly Information Unit (Acquisitions and Vendor Services) of Scholarly Information & Research (SIR) provides a range of statistics for this report (which differ from collected statistics for other reporting, eg. CAUL). Over the past year, questions have arisen with regard to accuracy/validity of statistics, particularly with regard to not only third party collated statistics, but also in-house statistics we collect for reporting. Ultimately it has led us to pose the question: are we comparing apples and oranges, or in reality bananas and gorillas?

The "Ideal" would be to be able to produce a one to two page summary of cost-per-download statistics where the facts and figures outweigh the footnotes. The "Reality" is that the spreadsheet produced has more footnotes than a Higher Degree Research thesis and in most cases still requires a certain amount of discussion (to explain the footnotes).

Today’s Presentation (Slide 3)

The complexity of usage statistics, whether they are gathered in-house or via third parties, has such a range that we can only scrape the surface in this presentation. We will endeavour to briefly outline our particular issues at Griffith University restricting ourselves to three specific topics: in-house collected statistics, statistics from third parties, and cost-per-download statistics that we use for Staircase reports. We intend to show through this presentation that usage statistics should be considered as one piece of information that adds up to the cost benefit of a resource.

However, we do not simply want to dump information upon you, we would like at the end of this session to get some sort of consensus from you on what is “the gorilla in the room”. We will summarise the Top 5 Issues with usage statistics for us and provide you with details and links to a short online survey to provide feedback.

In-house Statistics (Slide 4)

a. Reporting for Staircase Reports

Database usage is one of the metrics reported to senior management on a quarterly basis and confirmed annually. Often the statistics are not yet available for the quarter we are reporting on, as there is a lag between when the resource is accessed, and data is finalised and available. This results in an inefficient use of staff time as the usage reports must be recalculated again at a later date. For example, while metrics for Quarter 4 2012 were being gathered for Books 24x7, it
was noticed that there were discrepancies for Quarter 3 because these statistics were gathered by us in October (rather than November when they were fully released).

Griffith University subscribes to a number of databases that are not primarily text-based resources. Users may access images, video or audio content, and are not necessarily downloading full text content. The current Counter standard uses JR3 reports (Item requests and turnaways by Journal Title, Platform, Page Type and Month) for this data, however as this is an optional report and not all third party providers or publishers are able to provide the data (Sugarman, Krueger, Kelly, 2011). It is left to the discretion of the provider. This makes it extremely difficult to ensure we are measuring usage consistently. We subscribe to collections with two multimedia providers (Alexander Street Press and Adam Matthew Digital) and we currently receive DB1 statistics from both (eg. searches and session by month and database), which does not give the granularity of what exact resources are being used (eg. videos in the case of Alexander St Press and images in the case of Adam Matthew Digital).

Some providers do not produce any statistics. For example, we have a number of Digital Image Collections through LUNA Imaging, which include both LUNA collections and Griffith University images from our digitised slide collection from Queensland College of Art. LUNA is a worldwide network of hundreds of thousands of high-resolution, high quality digital and multimedia images, however, there are no usage statistics produced by the provider.

b. Reporting on ebooks (Slide 5)

The uptake of ebooks is steadily growing (which is reflected in the purchasing trends of ebooks by institutions in 2011 from this QULOC chart), but as usage can vary by discipline/subject area it is important that we understand and interpret the usage data (Cox, 2007). The Counter standards allow for 6 types of usage reports for ebooks, but not all publishers are able to provide this information. Some are able to provide downloads only, some searches and downloads per chapter and/or the entire work, and others by platform. The information provided is not always consistent and not helpful when looking at cost benefit. For example, at Griffith despite a significant increase in new ebooks purchased in 2012 (98,000 added to catalogue in 2012) there are still ebooks published that do not provide usage statistics. For example, the EOLSS (Encyclopedia of Life Support Systems) which is a compendium of 16 encyclopedias produced by UNESCO, and the Encyclopedia of Pharmaceutical Technology by Informa Healthcare.

Third Party Statistics (Slide 6)

At Griffith we use two third party services (360 Counter and Scholarly Statistics) to collect and compile Counter-compliant usage statistics. On examination, we identified the following inconsistencies:

a. Multiple means of access to platforms
Access via discovery layers, aggregators, etc. can have a direct effect on which resources are given a higher relevance. According to research completed in this area (Way, 2010, p. 219) libraries must "find a way to reconnect users to these resources" if they are still valued. At Griffith, there has been a decline in usage of content published by medium sized publishers, and Emerald has identified this trend, specifically since the introduction of Summon.

b. What constitutes a journal

There has to be some potential for inconsistency when a database provides JR1 statistics when the resources supplied are mainly ebook reports and ebook chapters, eg. Water Intelligence Online has access to ebooks and e reference books but counted as JR1.

c. Changes in publishers and/or platforms

Interruption to access because a publisher upgrades to a new platform can result in a significant drop in statistics. For example, in 2012 one of the Griffith law databases had technical issues when upgrading to a new platform, which resulted in an interruption to access and lack of usage statistics. When a journal title transfers to another platform, archival access via the new vendor may not match, which may result in a gap of holdings. For example, a journal title through the aggregator ATyphon, was taken over by Cambridge. The Griffith holdings were from 2006, the Cambridge holdings were from 2008, which led to re-negotiations for this gap in access and the need to secure the statistics from the aggregator.

d. Changes from p to e

Usage reports for journals that have recently moved to electronic form will start from zero and will only grow once more content is added and sometimes there are publisher delays in access.

e. Design of platform effects usage

Certain platforms are designed so that you get access via a HTML page (which is counted as a JR1) and if they proceed to click on the PDF version, this is also counted as a JR1, which then constitutes two downloads. This may result in inflated statistics for that database when compared with a different platform designs.

f. Usage peaks and troughs

Should the statistics be adjusted for lack of access due to a technical issue and/or a spike in usage due to a compulsory reading for say an exam?

g. Types of articles (Slide 7)

There are various journals that have editorials, book reviews, and short articles and they may be counted as a JR1 article download, in other words given the same value, as a research article. There is no differentiation between the type of article that is being counted. For example, two articles in BMJ (British Medical Journal): In Brief article of two sentences from 20/11/12 on promoting fruit/vegetable consumption compared to a 7 year longitudinal study of the effect of fruit/vegetable consumption and its effect on preventing dementia by K. Ritchie from 5/8/2010.
h. **Other Anomalies (Slide 8):**

There are issues with third party products trying to “normalise” title variations and therefore reporting usage twice (eg. duplicate Arion report from JSTOR). There is also some issue with differences between quarterly reports that are run, when compared to consolidated reports via Pivot Link in 360 counter. For example, the advice given via Serials Solutions webinars when there are duplicated statistics in the Consolidated Reports is to delete the Mid-Year Reports, however, this means each duplicate report needs to be deleted manually one report at a time. Therefore if for some reason there is more than one duplicate report, this can be a time-consuming process.

**Cost per Download Statistics (Slide 9)**

a. **Rolling front files**

With certain packages/collection there may be a rolling block of five years (eg. 2008-2012) which means in 2013 you will lose 2008 unless purchased via individual years or collections. Statistically it is difficult because if we did not purchase the archive to dovetail with the beginning of the subscription, we will end up with gaps and inaccurate comparisons for cost-per-download.

b. **Full-text content aggregator usage statistics**

Griffith has separate prices for various individual databases on the one platform. However, the JR1 figures are only generated over the whole platform. So we can arrive at a cost-per-download figure across the set in a platform, but not for each of these databases (unless we manually calculate separate figures).

c. **Backfile statistics**

There are arguments for and against deducting backfile statistics (JR1a). At Griffith we deduct JR1a statistics (and indicate this clearly) as many of our backfiles are paid for as a one-off purchase and not a subscription.

d. **Exchange rate fluctuations (Slide 10)**

In 2012, the AUD/USD currency exchange rates fluctuated 10% from a high of AUD1.07 to a low of AUD0.97 (dark blue on chart) and as you can see 2011 was even more volatile (light blue on chart). Whether the cost information comes from what we pay or the publisher price there are inherent inaccuracies. For example, if Griffith pays an estimated invoice as at a certain date (with its corresponding exchange rate) and then when the actual invoice is produced a credit or further charge is issued (with its corresponding exchange rate); the exact cost of the resource may be difficult to derive. There is also a question as to what date invoices are paid, who actually pays the invoice and what system processes the payment. If there is a third party, ie. Finance department, actually processing payment there may be issues with minor differences associated with exchange rates. This may also be exacerbated if the finance system and the Library Management System are not interoperable.
Summary (Slide 11)

Griffith is taking the view that a yearly set of statistics does not tell the full story. Often a single year of statistics on a single ebook/ejournal/database is merely a snapshot in time may not account for the anomalies we have outlined. 2-3 years of the same statistics produces the kind of trend data to make informed decisions (Katsouros quote from Educause Review, Jan 2013). However, the value of a database, journal or ebook also needs be judged within a larger context. The other important part is the strategic intent of the institution, which are comprised of the research, learning, teaching and student experience goals of the university. At Griffith, if a Research Centre funded specifically as a strategic research strategy is highly cited within a specific ejournal, then low usage data for it may not necessarily be the most important factor used to judge its value. For example, there was a deal of hand wringing with regard to China Academic Journal database with regard to low usage statistics when it was first subscribed to, now researchers are praising the Griffith library for making such a forward thinking decision and subscribing from 2006.

Griffith’s Top 5 issues with usage statistics (Slide 12):

So in order to follow the theme for this conference “Be different, Do different” we would like to invite you to participate in an online survey to see if the Top 5 usage statistics issues at Griffith University are important to you.

The top 5 usage statistics issues that we identified are (Slide 13):

1. Lack of more granular statistics
2. Lack of consistency with the ebook usage statistics provided
3. The implementation of discovery layers has skewed the figures
4. The delay in third party providers collecting/making the statistics available
5. Non-Counter Compliant statistics

We would like you to choose if each of these issues are either:

a) Banana sized issues, eg. small problem but usually has a workaround
b) Gorilla sized issues, eg. medium sized problem that requires extra manual work
c) King Kong sized issues, eg. big problem that ought to be pursued through our advocacy via official channels.

Question number 6 asks you to rank these issues in order of importance and question number 7 gives you the opportunity to outline a usage statistics issue that you would like addressed and why. There are 500 printed postcards to hand out with this information. Are there any questions (Slide 14)

Special Thanks to:

- Debby Macdonald, my co-writer, chief editor and silent presenter
- Anne Fleet from Acquisitions who deals with statistics on a daily basis
• Tara Anderson, Marketing & Communications Advisor, INS Communication Team for the fantastic postcard handouts
• Christine Gordon from James Cook University for her advice

Thank you and we look forward to your feedback by 31<sup>st</sup> March, 2013 (Slide 15)