

Big Deal and Long Tail: A Case Study on E-Journal Usage and Subscriptions

Joachim Schöpfel & Claire Leduc

Abstract

Purpose – This paper is aimed primarily at academic library managers and acquisition librarians. By analogy with Pareto's study of the relationship between clients and turnover, we will study subscriptions to e-journals and usage statistics. Our purpose is to evaluate the long tail of usage statistics and to compare it with the two modes of subscription, individually selected journals vs. packages (big deals).

Design/methodology/approach – The paper exploits usage statistics and subscription data from a national usage study of an academic publisher. Data are from 2010.

Findings – Usage statistics are partly shaped by the long tail effect. Individual subscriptions to journals are more selective than big deals, and tend towards a traditional retail model. Unlike subscriptions through packages, usage and individual subscriptions can be related by a similar inclination. But both types of subscriptions fail to predict the popularity of a journal in its usage.

Research limitations/implications – The paper uses data from a national usage study and tries to identify global trends. Thus, it does not distinguish between customer categories, disciplines and activity domains.

Practical implications – The paper considers the opportunity provided by big deal for acquisition policy. Ready-made big deals sometimes appear as an unbounded and excessive supply not suited to the users' true and sufficient needs, while selective acquisition policy cannot completely anticipate online usage behaviour.

Originality/value – Only a few studies distinguish Pareto from long tail distributions in usage statistics, and there is little empirical evidence as to the impact of selected subscriptions vs. big deals on these statistics.

Introduction

Since Chris Anderson popularized the long tail in 2004, the concept has challenged the academic library. On the basis of studies on Amazon and other e-commerce companies, Anderson suggested a new competitive business model described as a new niche strategy "selling less from more" (see also Anderson, 2006). The underlying principles consist in making everything available at a (very) low price and in driving demand down the long tail, through recommendations, filtering and other helpful tools. Recently, Peltier and Moreau (2011) illustrated the long tail effect with empirical data from the French book market.

The Anderson long tail distribution is a special variant of the so-called Pareto principle, also known as the 80-20 rule. This statistical principle applies in many disciplines and also in library and information sciences. It states that often 80% of the effects come from 20% of the causes. Here are some examples. In a given population, 80% of the income is controlled by the riche 20%. In a company, 80% of the revenue comes from 20% of the customers or products. In a library, 80% of the book-lending is done by 20% of the readers.

In the case of Anderson's long tail, this 80-20 rule moves to 80-30 or 80-50 or even more. For instance, in an online music store the most selling 20% disks no longer generate 80% of the revenues but less (70 to 50%, see Figure 1). In other words, 80% of the revenues are due to a larger number of products. This means also, that the "top of the charts" (bestsellers, most-read journals or books, highly requested and downloaded e-journals) become less important. The remaining less selling 80% disks are also called a "fat" or "heavy tail".

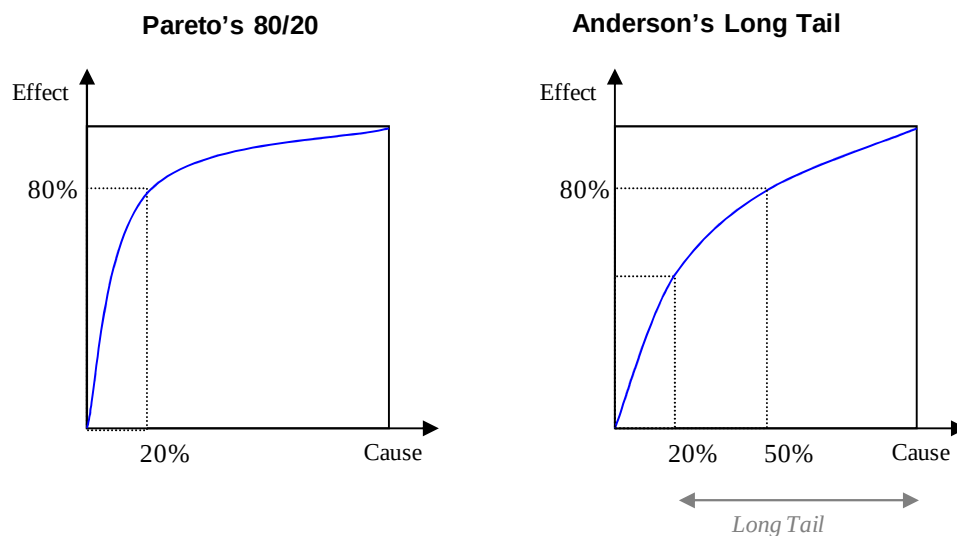


Figure 1: Examples of Pareto's 80-20 and Anderson's long tail distribution (here 80-50)

Librarians know the business model of an almost unlimited powerful search engine offering no selection or filtering by end users insofar as it corresponds to the rationale of the “big deal”¹. The big deal “in which publishers sell online subscriptions to large bundles of electronic journals” (Poynder, 2011) is different from traditional acquisition policy based on need-centred selection or on popularity/quality indicators such as Garfield's impact factor (see Thelwall, 2008).

The growing impact of big deals on academic libraries and communities cannot be denied. “Traditionally library sales were in the form of subscriptions to individual journals. This is still an important part (currently around half) of the market but increasingly journals are sold as bundles of titles, either directly to libraries or to library consortia” (Mark and Ware, 2009). Quite similar to log files from e-commerce, usage statistics of digital libraries reflect the reality of the long tail in library science. Even if digital collections do not delete “blockbusters”, “chart-toppers” or “bestsellers”, a significant part of downloads becomes long tail activity, e.g., access to low-used journals, backfiles or e-books that may not have become available to the libraries' patrons without big deals (Starr and Williams, 2008).

A paper on e-books usage revealed “small clusters of high-use material among low-use and no-use material (e.g.) an extreme version (...) of the power curve” (Christianson, 2005). Another study highlighted the importance of discovery services and concluded that “IT-enabled tools can enhance the effect of Long Tail markets by allowing consumers to actively and passively discover products that they otherwise would not have considered” (Brynjolfsson et al., 2006). In Anderson's terms, libraries started to move from physical and hybrid to pure digital retailers.

Academic publishers' selling point is that more content will lead to further and better usage, with positive impact on scientific quality and output. In addition, the growing interest of the long tail content may change library marketing and its relationship with its communities (see Ferguson and Hlavinka, 2006; Lee, 2007).

¹ See Storey (2005) or O'Beirne (2005) on the relevance of the long tail for libraries and Poynder (2011) for a thorough and critical review of the big deal business model.

A recent study showed the dynamics of this process in the field of document supply, with a decreasing demand for an increasing variety of items (Schöpfel and Gillet, 2007). Their data confirmed a change in the shape of the demand curve to what Elberse (2008) called a “fattened long tail”.

Now, in a new usage study, we had the opportunity to compare e-journal usage statistics with subscriptions to individual titles and big deal acquisition policy. The overall idea was to evaluate the reality of the long tail (low-usage journals) and to compare it with the subscription lists. Our question is this: Do big deal and long tail provide a new opportunity for selective acquisition policy?

In the following article, we report on global statistics on journal usage and subscriptions in order to assess the reality of the long tail effect and its relationship with acquisition policy. Our research hypotheses are:

H1: Online usage is shaped by the long tail effect.

The usual Pareto 80-20 distribution predicts that 80% of the usage (effects) would be related to 20% of the journals (causes) while the long tail theory assumes that 80% of the usage would be related to 50% of journals or more. We expect that in our study online usage is closer to Anderson’s 80-50 than to Pareto’s 80-20 distribution.

H2: Individual subscriptions are more selective than big deals.

We expect that individual subscriptions to print journals with access to the online version include fewer titles than subscriptions to the online versions through packages. We expect, too, that their distribution shows more similarities with the traditional 80-20 retail curve, with some “bestsellers” and a lot of journals with fewer or no subscriptions at all, while the big deal distribution of subscriptions is more similar to Anderson’s long tail curve.

H3: Usage trend is closely related to individual subscriptions trend.

It is likely that the long tail of online usage is globally more similar to the distribution of print subscriptions than to the big deal distribution. We assume that selective acquisition policy should in some way reflect knowledge of patrons’ needs. Therefore the distribution of selected subscriptions to print journals with online access should at least slightly correlate with their usage statistics.

H4: Yet, usage ranking differs from the order of individually selected journals.

We suppose that at least some journals with no or very few individual subscriptions are not part of the long tail but of the “short head” of popular journals with high usage. Our assumption is that selective acquisition policy cannot completely anticipate online usage behaviour and that for these items at least, a big deal approach appears to be a better solution.

Methodology

The empirical data on usage and subscriptions were provided by an international academic publisher. The sample consisted of 253 journals, all of which are available via the publisher’s online platform, and of 91 customers including several thousands of authorized end users (students, scholars, staff). The surveyed usage period was January to December 2010.

A total of 3,523 subscriptions and 71,094 online requests were analysed. The analysis of usage was based on the number of successful full-text article requests per month and per journal, as defined by the international COUNTER project; this being the only widely

accepted standard that “allows the usage of online information products and services to be measured in a credible, consistent and compatible way using vendor-generated data”.²

One part of the journals were subscribed individually, title by title, with acquisition of the print version and access to the online version on the publisher’s platform. Other subscriptions were part of big deals, e.g., access to journals on the publisher’s platform was licensed via packages (bundling). The publisher’s portfolio contains 25 multidisciplinary and thematic packages with 5 up to 212 journals (median size = 19 journals, average mean = 60 journals); no package covers the whole journal catalogue of the publisher.

The database linked the following information and datasets: some basic data on customers (type, sector, number of users) and information on subscriptions (type and number), some data on journals (subjects) and on packages (number of journals, subjects), and usage statistics (full-text article requests). The data were taken from different sources, i.e. from the publisher’s customer file, from the publisher’s catalogue (portfolio) and from a commercial database on periodicals for complementary information on the surveyed serials, and from the publisher’s weblog statistics produced by his platform. Export of the data, validation of the data and correction of incoherencies and errors, processing of missing data, import into the database took some months of work.

A small number of online requests that could not be matched to a subscription – due to a trial period or a perpetual access as a result of cancelled subscriptions - were removed from the database.

Results

The distribution of usage statistics for all journals (H1)

Online requests show the characteristics of a long tail. Figure 2 plots the journals on the horizontal axis (ranked from left to right by decreasing number of requests) and the usage on the vertical axis. Obviously, a few journals (on the left side of the curve) are viewed many times while many journals (on the right side of the curve) are less frequently accessed.

²See the COUNTER Code of Practice for e-Resources (release 4, April 2012) available at <http://www.projectcounter.org/r4/COPR4.pdf>

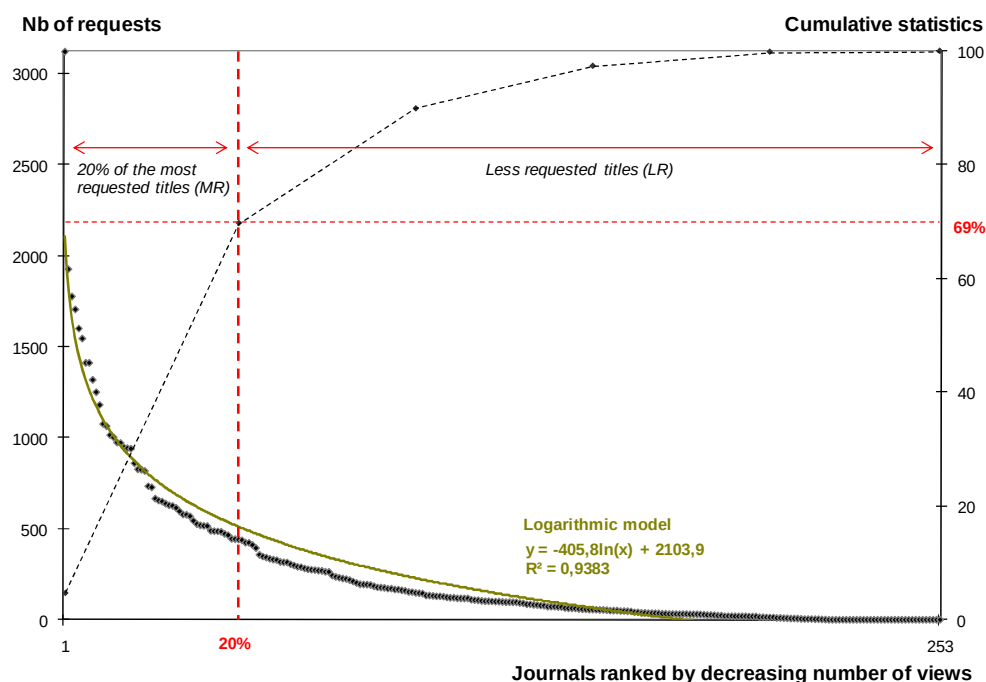


Figure 2: Usage of all journals: detailed and cumulative statistics

When cumulated, the 20% of the most requested journals (left side) do not sum up 80% of the usage (as expected by the Pareto principle) but only 69%; this is close to the long tail distribution but widely overpassing the expected limit of 50% assumed by Anderson's genuine long tail theory. In fact, the empirical curve of online requests is even closer to the usual Pareto distribution than to the Anderson assumption³. This means that our first hypothesis, *online usage is shaped by the long tail effect*, is only partly confirmed. The empirical reality of usage statistics seems somewhere in between the two distributions.

The distribution of individual subscriptions vs. big deals (H2)

As shown in Figure 3, most of the customers (74%) hold only subscriptions to individually selected print journals with access to the online versions. 7% of the customers have ordered journals via packages. The other 19% have ordered both, packages and individual titles, often regardless of double subscriptions (left pie chart).

³ For the confirmed reader: this distribution can well be adjusted by a logarithmic model (included in Figure 1).

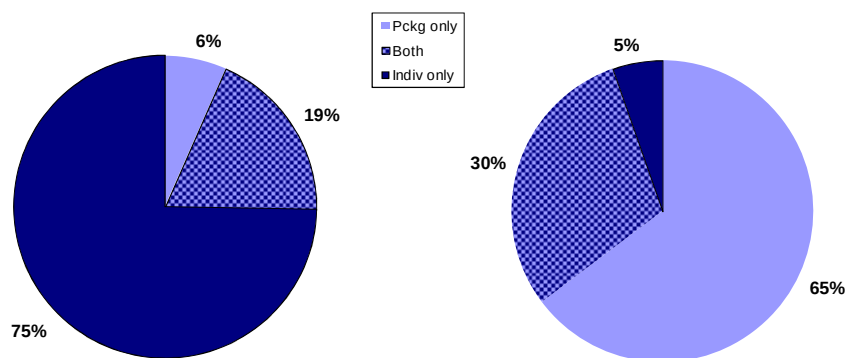


Figure 3: Subscriptions modes in terms of number of customers (left) and number of journals (right)

The pie chart on the right shows that these individual subscriptions only match to 5% of the overall number of subscribed journals, while the packages (7% of customers) cumulate 65% of the journals. This figure confirms if necessary the numerical importance of packages for the access to journals and articles.

Figure 4 shows for each journal the overall number of subscriptions, and distinguishes with different colours on the bars the two subscription modes (dark: individual subscriptions; light: subscriptions through packages).

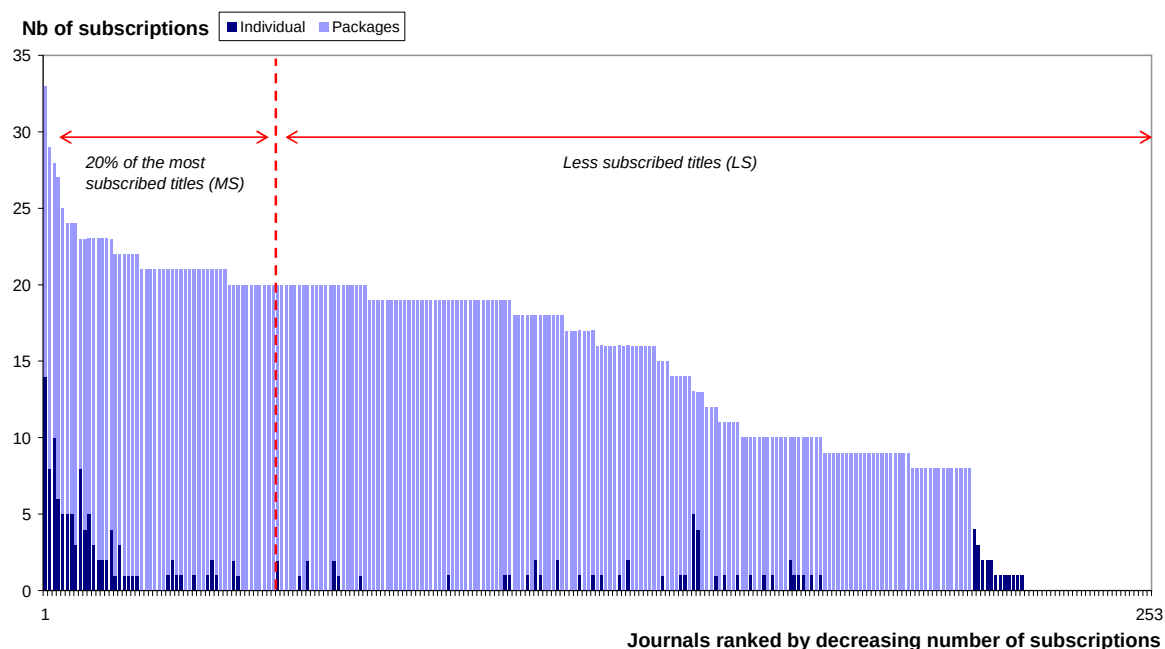


Figure 4: Overall subscriptions to journals

It appears clearly that subscriptions to individual journals represent only a small part of the overall subscriptions (5%). They also cover a smaller set of journals, only 31% of the samples (79 journals), compared to subscriptions via packages that represent 84% of the samples (212

journals). Subscriptions to individual journals focus on the same “bestsellers” as packages (left side of distribution, dark and light bars), but also target some niche journals not included in packages (right side of distribution, only dark bars).

Obviously, the empirical data confirm our second hypothesis, e.g. *individual subscriptions are more selective than big deals (packages)*. Individual subscriptions include fewer journals than subscriptions through packages, and focus sometimes on specialized journals.

If we study each mode of subscription separately, it appears, as expected, that the distribution of the individual subscriptions shows similarities with the traditional retail curve with some bestselling journals and a lot of journals with fewer or no subscriptions at all (Figure 5).

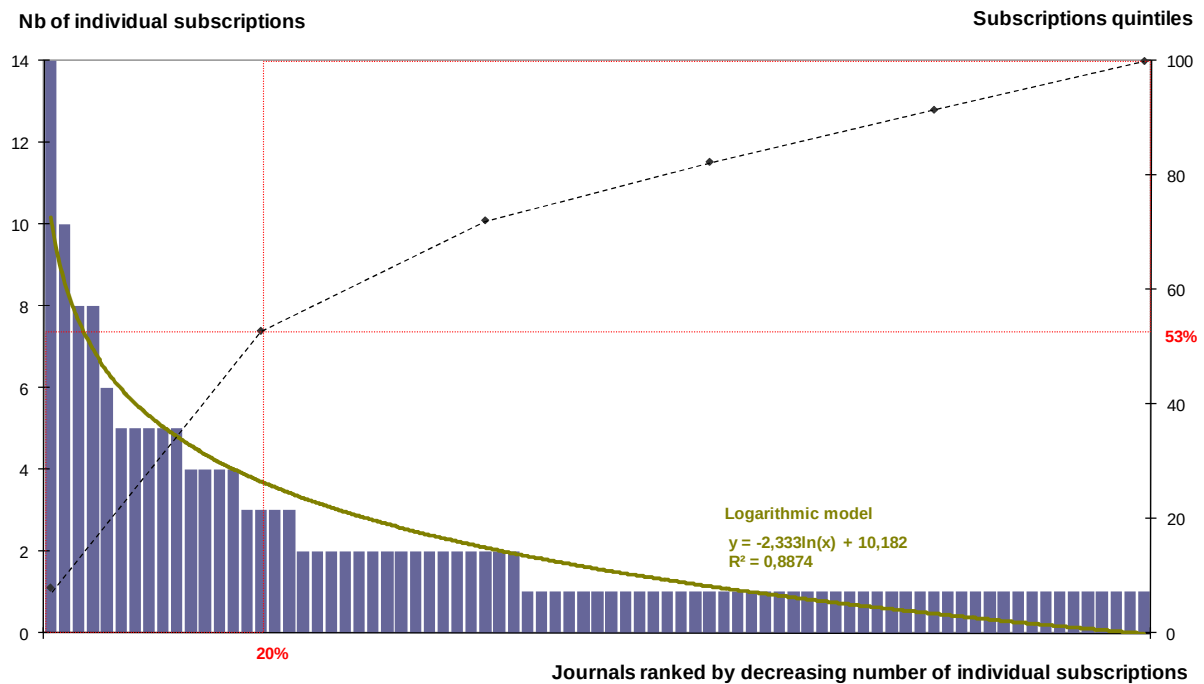


Figure 5: Distribution of individually subscribed journals (un-subscribed journals are omitted)

For this subscription mode, 20% of the journals (causes) correspond to 53% of the overall number of subscriptions (effects). This obviously is a long tail distribution, with some bestsellers of less importance.

On the other hand, the big deal distribution is not similar at all to a long tail curve (Figure 6). Its shape is concave and shows stages related to the number of journals inside each bundle.

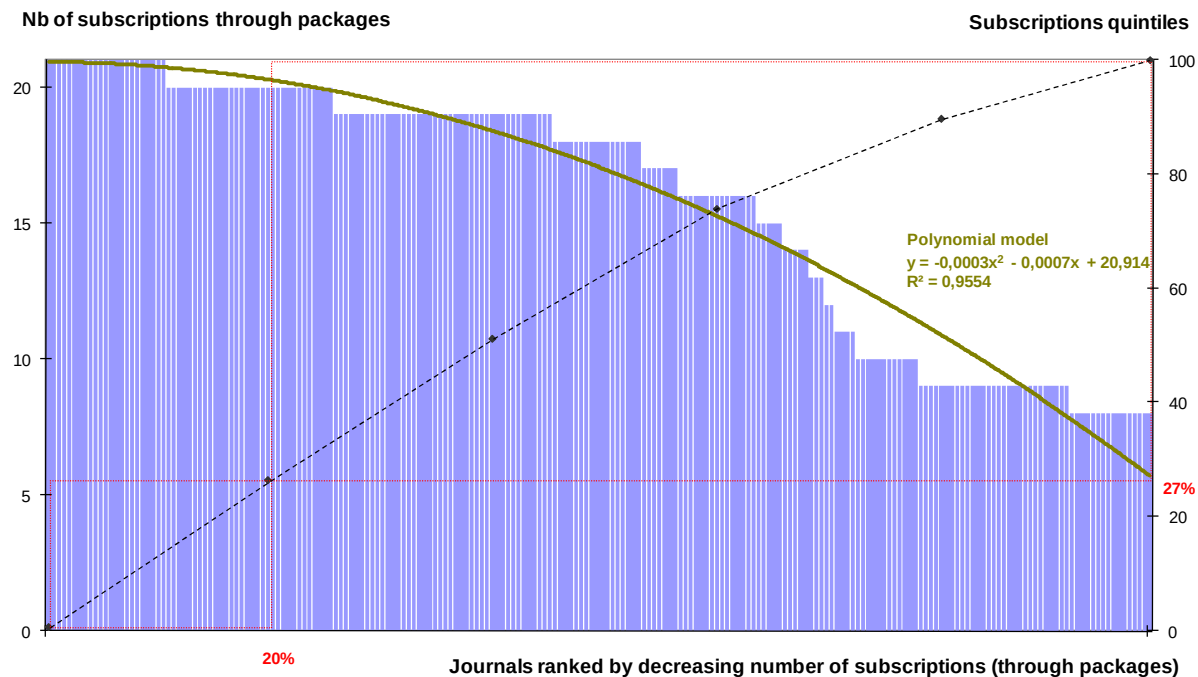


Figure 6: Distribution of subscriptions via packages (un-subscribed journals are omitted)

Here, the 20% most subscribed journals correspond to only 27% of all subscriptions while 80% of the journals correspond to 73% of the subscriptions. This is very different from the Pareto 80-20 rule or Anderson's long tail distribution of 80-50. One could even speak of an extended or inflated long tail, a rather interesting distribution of a commercial activity.

Rank correlation and similarities between subscription and usage (H3)

Now let us consider the relationship between subscription and usage. The following scatter-plot (Figure 7) matches the number of subscriptions (customer accounts) and full-text article requests for each journal. Each point represents a journal title.

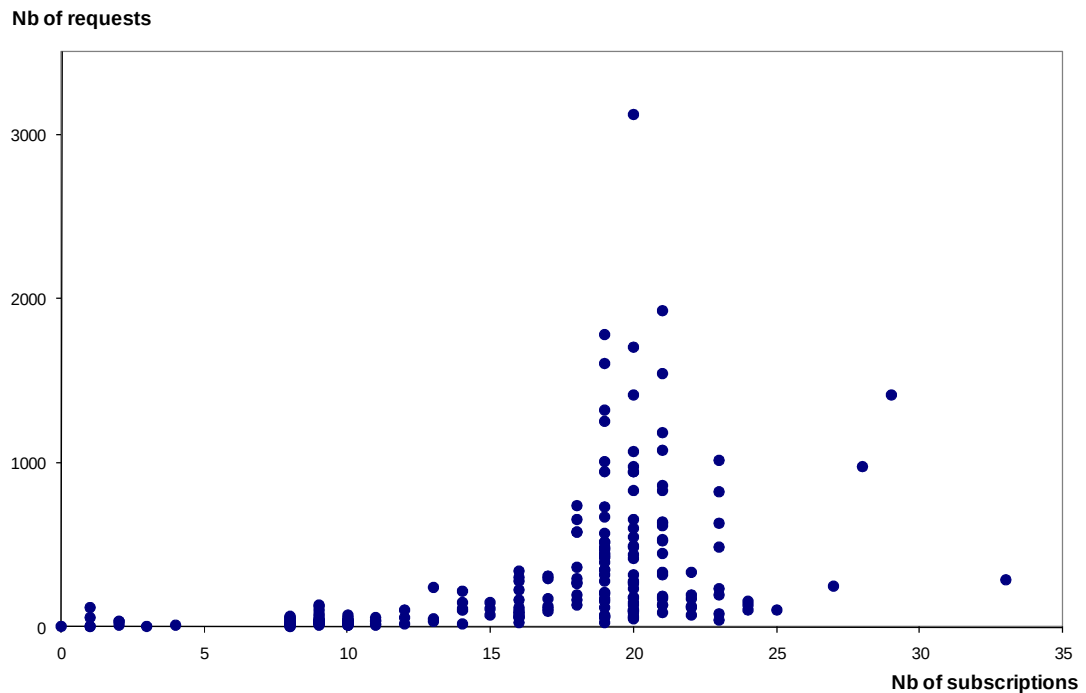


Figure 7: Subscriptions and usage statistics for the whole set of journals

The diagram shows that subscriptions and usage are not independent one from the other. The higher a journal ranks in subscriptions, the higher it will probably rank in usage statistics. Kendall's tau coefficient, a reliable statistic for rank correlations, is $\tau=.63$ and close enough to 1.00 to confirm this relationship. This high correlation exists also for the two specific modes of subscription and their corresponding usage, i.e. individual subscriptions and packages.

This leads us to go deeper into the analysis and to link usage with the two different modes of subscriptions. As we have already mentioned above, usage statistics and individual subscriptions are distributed in a way similar to the Pareto principle (80-20 rule), while the package subscriptions show quite a different distribution, what we called an “inflated long tail” (Figures 2, 5 and 6).⁴

This observation confirms our third hypothesis, that the distribution of online usage has more similarity with the individual subscriptions curve than with the big deal distribution, e.g. the *usage trend is closely related to the individual subscriptions trend*. As we expected, the “long tail” of online usage (= less requested journals) is globally more similar to the journal list with low or no individual subscriptions (= less subscribed or non selected journals) than it is to the big deal distribution where a lot of non selected journal titles are included in the subscribed packages.

The limits of individual subscriptions (H4)

Finally, we tried to appreciate the popularity of the journals with respect to their subscription type. We focused on the 20% most requested thus popular journals, which defined the short head of Figure 2 where they were referred as MR (for “most requested”). What was the major

⁴For the reader interested in statistics: We can adjust both the usage distribution and the individual subscription by logarithmic models, whereas the package subscription distribution is better adjusted by a polynomial model. Moreover, both logarithmic models can be linked by a geometrical transformation (a translation followed by a linear homothetic transformation, and then another translation) which proves the proportional relationship between both models.

type of subscription for these journals? Unsurprisingly, packages: all of them were subscribed through packages except for 27%, which were via individual and selected subscriptions. Let us now restrict our study to the 51 journals that were subscribed through packages only without any selection per title. How many subscriptions do these journals attract? Testing our H2 (see above), we distinguished the less subscribed journals (LS) from the most subscribed journals (MS) (see Figure 4). Here in Figure 8, we plotted the usage of the 51 package-only journals and coloured the LS titles in yellow. It is noticeable that only half of the 51 journals are highly subscribed journals (MS) while the others are part of the less subscribed journals (LS). We can conclude then that subscriptions through packages allow an unexpected popularity for many journals (reverse order compared to subscriptions).

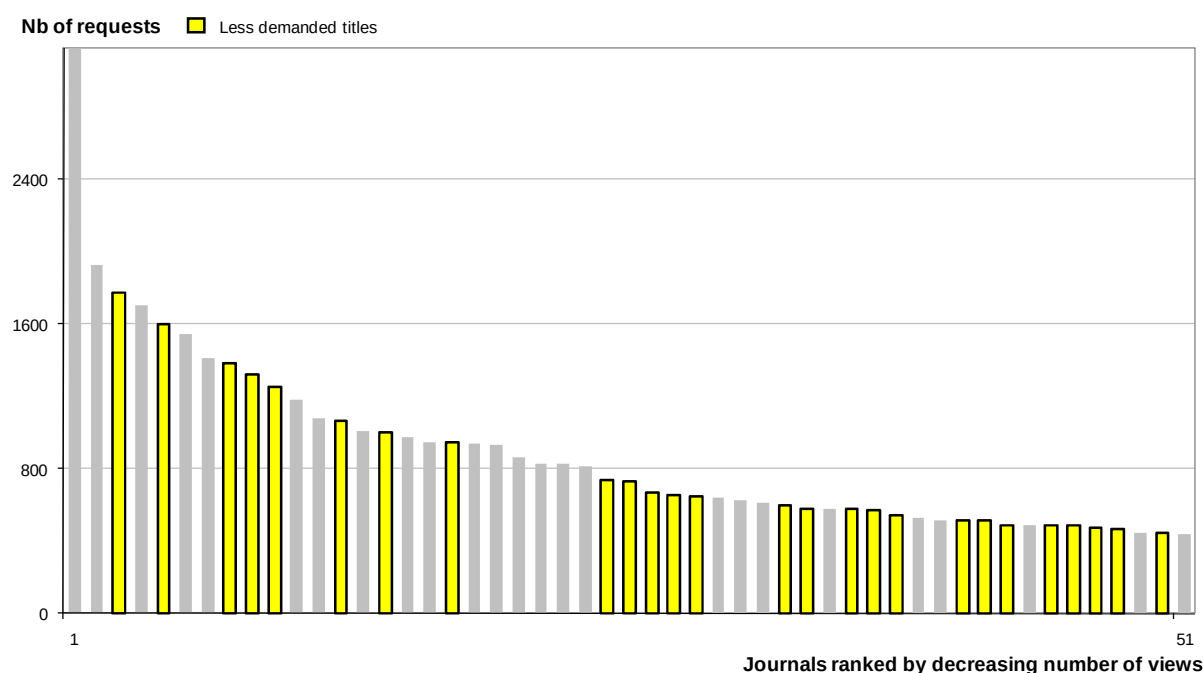


Figure 8: Usage of the most requested (popular) journals issued from big deals

In the same manner, we could focus on individually subscribed journals that are part of this short head of most popular journals with high usage (MR). 21% are part of the less subscribed journals (LS). These individual subscriptions correspond to .02% of the overall usage and to .03% of the “short head” usage.

The usage distribution is thus different from individual subscriptions and confirms our fourth hypothesis, that *usage ranking differs from the order of individually selected journals*. The selective acquisition policy cannot completely anticipate online usage behaviour and at least for these items, a big deal business model is a better solution.

Discussion

With regards to needs and usage, is a selective acquisition policy better than subscription to larger bundles of non-selected journals? Our results provide a contrasted answer.

First of all, authorized users of libraries and organizations that opted for big deals had access to a much larger offer than the others (H2). The high correlation between the number of subscriptions and usage statistics shows that the supply responded to the demand (H3). It seems obvious, too, that among the journals subscribed exclusively through packages (without individual print subscriptions) are highly used items (H4). The richness of the licensed

content and the real usage of journal titles that otherwise would not have been selected are two major arguments in favour of big deals. Our study contains elements that underscore these arguments.

On the other hand, the distribution of usage has more similarities with selected subscriptions (H3). In fact, the usage distribution still remains more similar to the traditional 80-20 model than to the expected long tail distribution for items online (H1). Figure 9 illustrates this observation.

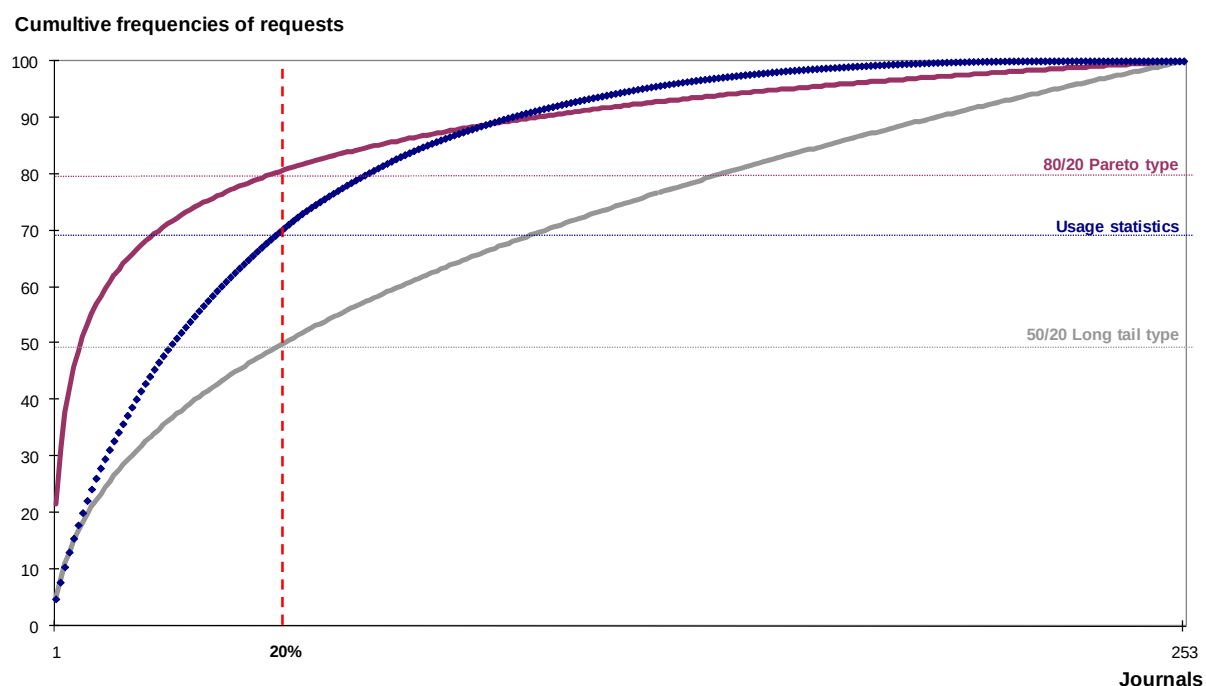


Figure 9: Cumulative usage of journals, together with ideal Pareto and long tail distributions

On the left side of the distribution (short head), the usage data follow the expected long tail distribution described by Anderson (50-20) but then progressively join the more classic Pareto distribution (80-20). This means that we can observe fewer highly accessed journals than in traditional usage distributions of print materials, especially for the first 30-40% of the publisher's portfolio. But so far we cannot observe the long tail of less used titles predicted for online traffic. This may express a transitional state of information behaviour, a situation somewhere between traditional library behaviour and e-commerce. But it may also be related to a lack of visibility and search facilities of the publisher's offer, which are conditions for the benefit of the long tail distribution. Authorized end users may not be totally aware of the richness of the licensed content and therefore fail to use the whole potential to which they have access.

Translated into acquisition policy, our argument is ambiguous. Licensing of packages responds to the users' needs insofar as they can find the content they are looking for and even more, thanks to the big deal. Nevertheless, there are lots of not accessed journals: extent of usage falls around 60% for some packages, and this has nothing to do with the size of the package. On the other hand, other journals are requested that are not part of big deals but were subscribed individually as print subscriptions with online access. Does this mean that packages are useless and can be replaced by individually subscribed "bestseller" journals? Again, the empirical evidence is more contrasted.

In our last diagram (Figure 10), we matched overall subscriptions and online requests for all journals, and applied the ABC-analysis. This analysis extends the Pareto principle from two portions (the 20% hits and the other 80% items), to three: (1) the 10-15% of journals that generate 60-70% usage; (2) the further 20% of journals that generate 30% usage; (3) the 65% of journals that generate 10% usage. This approach allows distinguishing very popular journals from lesser known and niche journals.

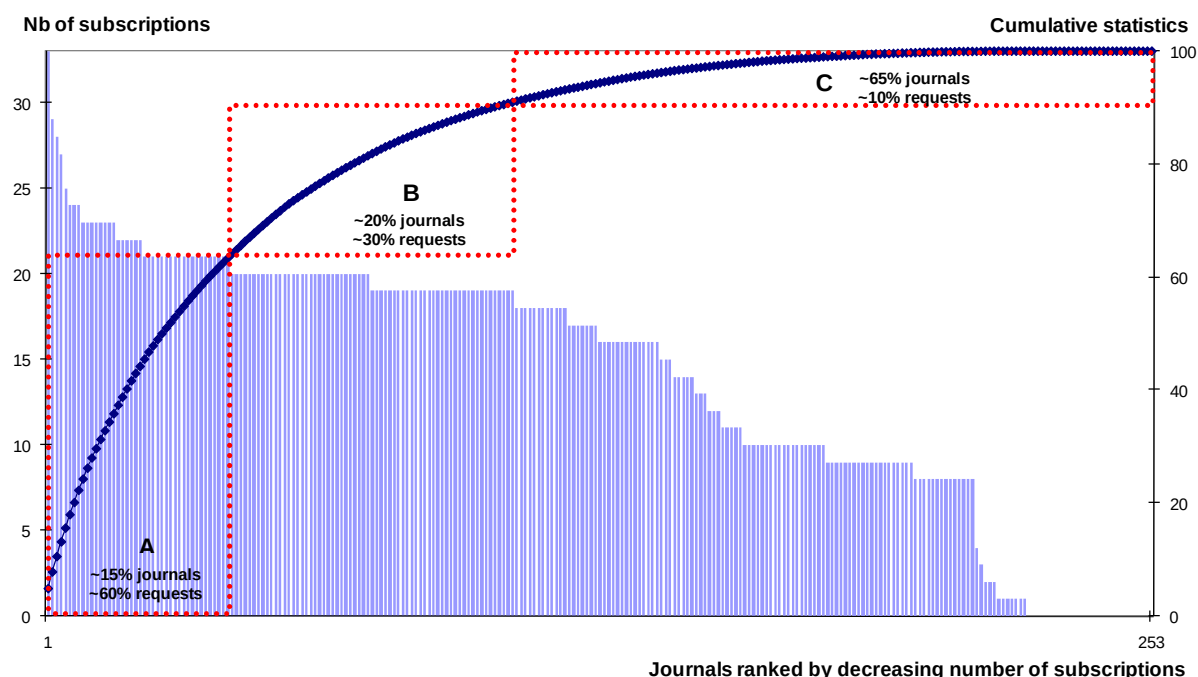


Figure 10: Subscriptions and cumulative statistics

Section A contains 15% journals that represent a bit more than 60% of the overall usage; section B, 20% journals for almost 30% of the usage, and section C the remaining 65% journals with only 10% of usage. Other sections could be defined to obtain an even more differentiated distribution. This manner of procedure may be an alternative to the two traditional groups 80-20 and 20-80 because it provides the publisher and customers with reliable arguments for the discussion on un-bundling and/or re-bundling of packages.

Cancellation of subscriptions, cutting the long tail for financial or even political reasons is a solution neither for the market nor for the end user who is in demand for content. Re-shaping big deals based on informed empirical evidence of usage and subscriptions may be an interesting alternative to explore. Several marketing models for subscriptions can be found in other domains, which give an idea of the wideness and flexibility of possibilities for publisher “ready-to-read” packages. As an example, one can subscribe to a “ready-to-view” theatre plays package which includes a fixed number of theatre plays among which only a few are imposed.

Our study raises two other questions. The first one relates to the definition of the long tail. In theory, it covers 80% of the right side of the distribution, and that is the threshold we have retained. However, our study shows that this cut-off point may shift to another value because of the characteristics of the empirical data. Indeed, the bundling of journals creates a kind of artificial levels in the distribution of the demand, and it may appear unnatural to cut the distribution in the middle of such a section (see Figure 4).

The second question is about the impact on mediation. If academic or research libraries want to re-negotiate content and package licensing, they need to understand usage features and get hold of data. According to Ferguson and Hlavinka (2006) they should prefer detailed analysis of user/content and niche material to “mass marketing”. For similar reasons, Ball (2004) considers big deals as “bad deals” for universities and encourages academic libraries to pursue identification of users’ needs and specification of information sources.

The significance of our empirical results may be slightly reduced by some methodological shortfalls. We use data from a national usage study and we try to identify global trends. Thus, our study does not distinguish between customer categories (universities, business schools, corporate companies, governmental agencies...), nor do we separate different disciplinary or activity domains.

Another point that can be raised deals with the very nature itself of each mode of subscription. Big deal subscriptions allow online usage only. Individual subscription provides a print version of the journal. One third of individual subscriptions also appear inside a package (see Figure 3). This means that we not only discount the print usage (but that was not the subject of our study) but also the usage statistics amalgamate both types of subscriptions.

Another specificity of our data is related to the publisher’s portfolio. As we explained above, some of the packages contain a rather small number of journals and have more similarities with customized, made-to-measure offers than with big deals. Also, like other exploratory usage studies, our analysis is limited to one publisher with a specific profile of content, customers and data.

Yet, we assume that these aspects do not cast any reasonable doubt on the global validity and significance of the reported results, and we are convinced that our results can contribute to the discussion on acquisition policy and business models in academic libraries.

Conclusion

Customers of digital collections are moving from physical and hybrid to pure digital “retailers”. In terms of usage and subscriptions, they come close to the long tail distribution but, as our data indicate, there is still some way to go. In the past, subscription decisions were often related to journals’ quality measures, in particular to Eugene Garfield’s impact factor. Today, decisions shift from high-demand and high-impact journals to big deals and long tail usage distributions. In other words and quoting two famous heroes from cartoons, we observe an evolution from *Garfield* to the long-tailed *Marsupilami*!

A good theory should predict empirical data. So far usage studies have most often confirmed the Pareto-distribution characteristics of usage statistics but failed to distinguish and analyse differences between these distributions. More empirical evidence is needed to understand the reality and dynamics of usage, with more and other publishers, more customers (portals, consortia...), and more products.

Is there a future for acquisition policy? According to our results the answer seems affirmative. There is a life of inter-mediation beyond big deals. Often, customers are happy with big deals because they find what they want. Yet, this is not enough. They ask for more, for journals and articles not available via big deals.

This opportunity for renewal of inter-mediation depends probably also on the capacity of libraries to test new business models (Ball, 2004) and to link acquisition policy to other questions, e.g. to join for instance the debate on academic publishing and open access (Poynder, 2011) and to define new roles in the emerging long tail environment, such as publishing of new journals (Gould, 2009), uncovering tail data in institutional repositories (Heidorn, 2008) etc. In 2011 and 2012, academics and librarians in France and other countries protest against some STM publishers’ business models of big deals and high price policy. This may raise consciousness and new interest in inter-mediation.

References

- Anderson, C. (2004), "The Long Tail", *Wired*, Vol. 12, No. 10.
- Anderson, C. (2006), *The long tail: why the future of business is selling less of more*, Hyperion, New York.
- Ball, D. (2004), "What's the "big deal", and why is it a bad deal for universities?", *Interlending & Document Supply*, Vol. 32, No. 2, pp. 117-125.
- Brynjolfsson, E., Hu, Y.J. and Smith, M.D. (2006), "From Niches to Riches: Anatomy of the Long Tail", *Sloan Management Review*, Vol. 47, No. 4, pp. 67-71.
- Christianson, M. (2005), "Patterns of use of electronic books", *Library Collections, Acquisitions, and Technical Services*, Vol. 29, No. 4, pp. 351-363.
- Elberse, A. (2008), "Should You Invest in the Long Tail?", *Harvard Business Review*, Vol. 86, No. 7/8, pp. 88-96.
- Ferguson, R. and Hlavinka, K. (2006), "The long tail of loyalty: how personalized dialogue and customized rewards will change marketing forever", *Journal of Consumer Marketing*, Vol. 23, No. 6, pp. 357-361.
- Gould, T. (2009), "The Future of Academic Publishing: Application of the Long-Tail Theory", *Publishing Research Quarterly*, Vol. 25, No. 4, pp. 232-245.
- Heidorn, P. B. (2008), "Shedding Light on the Dark Data in the Long Tail of Science", *Library Trends*, Vol. 57, No. 2, pp. 280-299.
- Lee, D. (2007), "The Long Tail Implications for Marketing Library Services", *Library Leadership and Management*, Vol. 21, No. 2, pp. 93-94.
- O'Beirne, R. (2005), "The long tail in libraries", *Reference Reviews*, Vol. 19, No. 5.
- Peltier, S. and Moreau, F. (2011), "Internet and the 'Long Tail versus superstar effect' debate: evidence from the French book market", *Applied Economics Letters*, Vol. 19, No. 8, pp. 711-715.
- Poynder, R. (2011), "The Big Deal: Not Price But Cost", *Information Today*, Vol. 28, No. 8.
- Schöpfel, J. and Gillet, J. (2007), "On document supply in the digital world", *Interlending & Document Supply*, Vol. 35, No. 4, pp. 195-204.
- Starr, S. and Williams, J. (2008), "The long tail: a usage analysis of pre-1993 print biomedical journal literature", *Journal of the Medical Library Association*, Vol. 96, No. 1, pp. 20-27.
- Storey, T. (2005), "The long tail and libraries", *OCLC Newsletter* (April/June), pp. 6-9.
- Thelwall, M. (2008), "Bibliometrics to webometrics", *Journal of Information Science*, Vol. 34, No. 4, pp. 605-621.
- Ware, M. and Mabe M. (2009), "The STM report: An overview of scientific and scholarly journal publishing", report, STM International Association of Scientific, Technical and Medical Publishers, Oxford.

Acknowledgments

Acknowledgments to Nicolas Barts, Cherifa Boukacem, Audrey Knauf, Thierry Lafouge and Helene Prost for their contribution to the usage study and/or for helpful advice for the interpretation of statistics, and to the Emerald Publishing team for providing usage data.

Biographical Details

Joachim Schöpfel is senior lecturer at the department of information and document sciences at the Charles de Gaulle University of Lille 3 and researcher at the GERiCO laboratory. He is also director of the French national reproduction centre for PhD theses (ANRT). He is interested in scientific information, academic publishing, open repositories, grey literature and usage statistics. He is member of GreyNet and euroCRIS.

Claire Leduc is senior lecturer on computer sciences at ISCID - University of the Littoral and the Cote d'Opale at Dunkirk. She received a PhD in computer science from the University of Provence at Marseille in 1997. She published on geometrical modelling and scientific data visualization. Her research interests include data mining, information visualization, multivariate visualization and academic online publishing.