

The Long Tail(s) of the Law: An exploratory study

Graham Greenleaf, Philip Chung and Andrew Mowbray

What is the 'long tail'?

Internet commerce has invigorated interest in a branch of economic theory commonly described as 'long tail' theory:

The Long Tail or long tail refers to the statistical property that a larger share of population rests within the tail of a probability distribution than observed under a 'normal' or Gaussian distribution. The term has gained popularity in recent times as a retailing concept describing the niche strategy of selling a large number of unique items in relatively small quantities – usually in addition to selling fewer popular items in large quantities. (Wikipedia entry 'Long Tail')

Its most prominent exponent is Chris Anderson (*The Long Tail: How endless choice is creating unlimited demand*, Random House, 2006). 'The secret to creating a thriving Long Tail business can be summarised in two imperatives: (i) make everything available; (ii) help me find it.' (Anderson, 217). In Internet commerce (Amazon, eBay etc), the key elements in long tail economics are the replacement of a finite and partial inventory dictated by shelf space (eg in a music store) with the near infinite inventory made possible by Internet distribution (eg iTunes); the reduction of transaction costs; and good search facilities (often accompanied by recommendation facilities). Where the right conditions apply, the majority of demand for content shifts from the head of the sales volume/content distribution curve (the 'hit parade') to less popular items and tends to continue with some small level of sales being recorded for virtually all items in the inventory (ie the long tail). Although unlistenable to most people, most musical items turn out to be something that someone with obscure tastes wants to hear again. Provided the transactions (and inventory) costs are low enough, all sales in the long tail can also turn out to be profitable.

The theory of the long tail overlaps a number of other significant theories including:

- (i) the economics of abundance (as opposed to scarcity): what economic principles apply to a resource that is abundant rather than scarce?
- (ii) theories of commons, and the rebuttal of the 'tragedy of the commons' by shared digital goods.

Anderson proposes 'nine rules of successful long tail aggregators', which can be paraphrased as follows:

- 1 Reduce inventory costs: digital inventory is the cheapest of all; eliminate atoms;
- 2 Let customers do the work: peer production and crowdsourcing; examples are user-generated reviews, and customer self-service.
- 3 One distribution method does not fit all: multiple distribution channels are necessary to reach more market niches.
- 4 One product doesn't fit all: 'microchunking' and re-combinations serve more market niches.
- 5 One price doesn't fit all: variable pricing according to (for example) immediacy or convenience.
- 6 Share information: lose control of information; eg allow users to rank items by different methods.
- 7 Think "and" not "or": offer alternatives;
- 8 Let the market do your job: don't try to predict the market by what you offer; put it all out there and let the market sort it out.
- 9 Understand the power of free: business models based on add-ons and premiums; free samples

All of Anderson's rules may well be relevant to free access to legal information, but it is the last that is the impetus for this paper: free access to goods is one driver of successful long tail models of sustainability.

What long tails might exist in law?

Long tail theories are not primarily about public goods (or commons), or as we tend to refer to it in the discipline of law, 'free access to law'. They also have implications for electronic commerce in legal information through the reduction of inventory costs. However, there are interesting overlaps intersections between the implications of long tail theories, and the development of commons (public goods).

This paper is an initial exploration of one of those intersections. It is not about economics, but rather about how what examples of 'long tail' behaviours may occur once legal information is available according to the two imperatives identified by Anderson: (i) complete collections not previously readily available to most lawyers are now freely and conveniently available ('make everything available'); and (ii) full text search engines, and other location techniques, make information more findable than traditional print indexes ('help me find it.').

Where might we find long tails in free access legal information? There seem to be two initial questions worth investigating in relation to case law:

- (i) Do accesses to case law exhibit long tail behaviour? If there is unlimited and convenient access to all case law, will accesses still concentrate on a relatively small number of very popular cases, or will users access a very wide variety of cases, with almost the whole spectrum of available cases receiving some usage?
- (ii) Do citations of case law exhibit long tail behaviour? Under these same conditions

Of course, we would also like to determine whether free and convenient access to an 'unlimited inventory' of case law actually changes the behaviour of users of legal information compared with more limited forms of access. Finding good 'before and after' data is a more difficult exercise.

It seems obvious that whatever behaviour is shown by usage patterns and citation patterns, they will not be independent of each other. Cases that are very frequently cited are likely to be accessed more often. Citators are not only one way of representing citation statistics, they are also one of the search/location tools that influence which cases are accessed.

Some initial examples

No one free access system holds all the case law information relevant to answering the above questions, including access data, even if the enquiry is limited to the common law world. It is therefore necessary to focus on specific sets of case law. Ideally, this should be done where (i) a system has comprehensive coverage of all cases; (ii) it provides the only significant free access location for those cases (so as to hold usage statistics); and (iii) a citator is available showing citation of those cases by at least most significant sources of such citations.

The decisions of the *Federal Court of Australia* represent one of the best sources of Australian data for this purpose. Since its inception in 1995, AustLII has provided all decisions of the Federal Court since its creation in 1977. During the 18 years prior to AustLII the complete set of these decisions was not readily available to most Australian lawyers (or others). Print publishers only published a small selection of cases. Access to the commercial online service carrying them (SCALE via InfoOne) attracted only a very small number of users because of price.

The two databases of Federal Court decisions (there is a separate database for Full Court decisions) are the largest database for any Court on AustLII, and have the highest annual access rates for any court. It should therefore be possible to determine whether these accesses exhibit long tail behaviour. AustLII's free access LawCite citator provides a near complete citation history of all Federal Court decisions by all Australian courts and tribunals, and significant coverage of overseas citations, so it should also be possible to determine whether the citation of Federal Court decisions exhibits long tail behaviour. Whether these behaviours change over time can also be determined.

Another potentially valuable example is the *English Reports 1220-1873*, containing much of the foundations of the common law. The 125,000 cases in the English Reports have been available on CommonLII since December 2008, and are the only free access version available. Prior to then (and continuing) they were available from Justis on a CD ROM, and otherwise were only available from the relatively few law libraries that had a complete set of the more than 100 print volumes. The LawCite citator became available at the same time as the English Reports.

With the English Reports, one question is whether only those cases which have been famous over the centuries receive high access rates? Or does the available evidence indicate that, since search engines have enabled more effective methods of access, large numbers of pre-1873 cases are still regarded as relevant enough to current problems for legal researchers to access them? In relation to accesses, we only have post-2009 statistics, but can at least attempt to answer this question for current free access usage. We can also examine, from all the data available to AustLII, what is the distribution of the 125,000 cases in terms of which

ones are cited in subsequent cases, over the whole period for which data is available, and from all countries from which it is available.

It would be good if we could look for any evidence that free Internet access to the English Reports correlates with any increase in citation of cases from the Reports (ie total numbers of citations), and whether more individual cases are now cited than was previously the case. However, pre-2009 and post-2009 initial comparisons may not yet be possible because of limitations of AustLII's current tools, and in any event two years is too brief a period to draw conclusions about effects on citation patterns.

This paper will provide what evidence is available from AustLII's access logs, and from the citation data it holds. To some extent it is a scoping exercise of the extent to which the data accessible to AustLII, and the tools to analyse it, can support this type of research. It will only be a very exploratory study, but will be a first step in a field that is yet little trodden.