Giant Pools of Content: Theorizing Aggregation in Online Media Distribution

by Andrew J. Bottomley

I t has been fifteen years since *Wired* magazine editor Chris Ander
son first introduced the concept of the "long tail" for describing
how the contemporary online media environment was reshaping
the production and consumption of media content and other cul
tural products.[1] In economic terms, the basic premise of the long tail
is that the internet lowers the barriers to entry found in traditional
media distribution, such as channel capacity constraints in radio and
television broadcasting, shelf space limitations in physical bricks and
mortar retail stores, and editorial gatekeepers. Online, any and all
content can be made available regardless of its origins or popularity,
and digital search technology and recommendation systems allow
audiences to discover niche and otherwise obscure content. Thus,
circa 2004, Anderson's long tail theory posited that our economy and
culture were shifting away from a model of mass media that focused
on a relatively small number of mainstream "hits" at the head of the
product demand curve and toward a huge number of niche products
and markets in the tail (hence, the "long tail"). It suggests that the
potential combined audience size for niche, low popularity content
may someday rival that of the large audiences for popular mass me
dia content. At its core, the concept of the long tail   much like Tim
O'Reilly's contemporaneous Web 2.0 concept   was a business model,
conceived of as a solution to the failures of web commerce during the
earlier dot com era (1995 2001).[2] Nevertheless, the long tail concept
carried deep cultural significance, promising to democratize and diver
sify the production and consumption of media.

Yet the niche marketing phenomenon that Anderson described in
2004 was hardly new. He was simply articulating a new way for online
businesses to strategize and monetize their offerings. The niche media
producers and content that populated the long tail always existed; they

---

1 Chris Anderson, "The Long Tail," *Wired*, October 1, 2004, https://www.wired.com/2004/10/tail.

2 Tim O'Reilly, "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," *O'Reilly Media*, September 30, 2005, https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

just were not economically advantageous in a mass media economy. Moreover, the means for distributing and accessing these cultural products online had been developing for a decade. In Anderson's original 2004 *Wired* essay, he points to digital platforms like Amazon, Netflix, Apple iTunes, and Rhapsody, some of which dated to the mid 1990s. These were online retailers and media subscription services that all operated off one central principle: content aggregation.

In this essay, I historicize the concept of aggregation by tracing the rise of the web portal, specifically highlighting audio media and the case of AudioNet (later Broadcast.com). AudioNet innovated the vertical portal idea for audio and video content in the late 1990s, signaling a shift toward aggregation as the primary distribution model in the digital media industries. By "content aggregation," I am referring to the practice of pulling media content from various sources and making it accessible at one dedicated, easy to find location. Nearly all the major digital media companies and platforms of the post 1990 web era operate through principles of aggregation: Google, YouTube, Netflix, Spotify, iTunes, and even social network sites like Facebook and Twitter. Aggregation lies at the heart of one of the great promises of the internet: the ability of individuals to access whatever they want, whenever and wherever they want.[3] Tied up in these utopian visions are metaphors of abundance, prosperity, and democratization.

Historically, modern computing and the internet are founded on the principle of abundance (and overabundance). Vannevar Bush's key observation, articulated in his prescient 1945 article "As We May Think," was that, although the scientific community was producing a trove of vital research and information, it was becoming impossible for researchers to sort through this immense maze of information and make practical use of it.[4] He argued that scientists had a responsibility to share their research more widely and to find ways to make knowledge more accessible, to give individuals more control over information. Bush's solution to these problems of information overabundance and inaccessibility was computer like machines, such as his theoretical Memex (or memory extender), which could enhance collaboration and perform functions such as compression, storage, and retrieval.[5] During the 1960s and 1970s, the internet followed this model; it was primarily a research oriented tool for scientists and other knowledge workers to share information and communicate with one another.[6] In other words, it served to aggregate both users and data; the internet and its simple, adaptable network protocols enabled a varied, geographically dispersed community of scientists, technologists, military personnel, and graduate students, not to mention a growing international cadre of computer

---

3   This is is, of course, a false promise. Many critics have called out this myth of on-demand "endless choice." See, e.g., Chuck Tryon, *On-Demand Culture: Digital Delivery and the Future of Movies* (New Brunswick, NJ: Rutgers University Press, 2013), 177.

4   Vannevar Bush, "As We May Think," *Atlantic Monthly*, July 1945, 101–108.

5   The Memex is a theoretical machine that was never built. It is essentially a desk-sized mechanical computer that combines elements of modern computers and the internet: it stores books, records, and communications, and is mechanized so that the researcher can consult these materials with speed and flexibility. Among its features is the ability for users to keep track of their research paths, akin to practices of bookmarking and associative linking.

6   Janet Abbate, *Inventing the Internet* (Cambridge, MA: MIT Press, 1999), 113–145.

hobbyists, to virtually assemble and to share ideas that would otherwise have been isolated. Herein lies the source of the discursive rhetoric of access, openness, flex ibility, and decentralization that is so often attached to the internet, and the pre sumptions of democratization and diversification that flow from them.[7] Of course, as the media scholar Thomas Streeter's study of "internet romanticism" reveals, this popular perception of the internet as open and disruptive is myth, a product of late capitalist social and political thought and historical peculiarities.[8]

Although access to an abundance of people and information is the desired goal in our networked culture, much as Vannevar Bush envisioned, this access invariably brings with it a problem of overabundance that must be addressed through methods of sorting and selection.[9] Thus, aggregation online always involves elements of both pro fusion and control. The earliest web directories and search engines are prime examples of this seeming contradiction. Starting in 1992, as the US government opened up the internet for commercial use and Tim Berners Lee's World Wide Web system helped make it possible for organizations and individuals to create their own websites with (relative) ease, there was a profusion of new websites on the open web. Findability quickly became a looming issue: if you were a web user in the mid 1990s, there was plenty of content to engage with online, yet it could be difficult to locate, especially if you were looking to discover something brand new. Thus, the first web directories and search engines sprung up to help give users mediated access to the web. The earliest search engines, such as Archie (launched in 1990), were essentially indexes of public file listings, helping users find files (e.g., texts, images, software) that were scattered all over a vast network. Web directories, such as the original Yahoo! (launched in 1994), similarly presented a list or catalog of links to websites. Much like a mail order shop ping catalog makes it possible to browse a huge array of consumer goods, these web directories gathered up hundreds and even thousands of hyperlinks into one location, giving users a perusable inventory of the web.

These web directories were often organized by specialized categories, geographic regions, and languages   again, like shopping catalogs. For example, there were web directories devoted to internet radio stations, such as the MIT List of Radio Stations on the internet. Rudimentary in design, the directory simply provided a listing of radio station websites with hyperlinks, plus limited details about each station's geographic location and programming format. This made it easy for radio listeners to find out whether they could listen to their local radio stations online and to discover new radio stations or listen to distant stations otherwise unavailable to them over the terrestrial AM/FM airwaves. Functioning as an aggregator, a directory like the MIT List helped

---

7   The irony of this rhetoric of openness and democratized access should hopefully be clear. While this collection of scientists, technologists, and hobbyists may have seemed like a grassroots community in comparison to the hierarchical bureaucracies of the US military or IBM, access and control were nonetheless centralized among an (almost entirely white, male) educated elite. There were, and still are, significant barriers of access keeping out minority voices.

8   Thomas Streeter, *The Net Effect: Romanticism, Capitalism, and the Internet* (New York: New York University Press, 2011), 169.

9   Today, this problem is frequently addressed through the concepts of "information overload" and the "attention economy."

make internet radio content findable online, yet it also brought a top layer of control to the media ecosystem.[10] Web directories and search engines inherently perform acts of curating, reorganizing and representing the content of the web for their users.[11] As users become reliant on these aggregators to access the web, any website or content that is relegated to a lower status or left out of the aggregator's rankings entirely be comes practically unfindable. Circa the mid 1990s, if an internet radio station was left off the MIT List, it was almost as if that website did not exist.

As the number of websites and users exploded in the 1990s, directories transitioned to portals, and the curatorial focus became more targeted and medium specific. Yahoo! expanded from a human edited web directory in the mid 1990s   originated under the name "Jerry and David's Guide to the World Wide Web"   to a major web portal in the late 1990s. Indeed, the late 1990s and early 2000s were the era of the web por tal: specialized aggregator websites that brought together diverse media and informa tion sources in one place.[12] The idea was that users needed a home base from which to navigate the web   and to this point, many portals were also search engines, such as AOL, Yahoo!, Lycos, and Excite. These were de facto gateways or front doors to the web, and many users made them their browser homepages. Some portals like Yahoo! brought together broad swaths of content: news, weather, entertainment, shopping, a little bit of everything. Others, like AudioNet, focused narrowly on content from a specific market or niche   these sites were known as vertical portals.

AudioNet may have had a limited life span from 1995 to 2002, yet the content ag gregation model it helped pioneer has largely come to define the shape of media dis tribution and consumption online. Better known as Broadcast.com (the name the site adopted in 1998, three years after its 1995 founding in Dallas, Texas), AudioNet was a streaming radio, music, and (later) television vertical portal, self described as "the leading aggregator and broadcaster of streaming media programming on the web."[13] AudioNet is a significant case study in the history of web media for a number of rea sons, not least because it brought its CEO Mark Cuban to wealth and fame. During its initial public offering in July 1998, the "frenzy" over Broadcast.com stock raised the company's valuation by more than $1 billion, resulting in what still remains the most profitable opening day gain of any company in Wall Street history.[14] Then, in March 1999, Broadcast.com was sold to internet giant Yahoo! in a stock swap deal worth $5.7 billion, which made it one of the most expensive transactions of the dot com era and Yahoo!'s highest priced acquisition of all time. The site was discontinued in

---

10 Alexander Galloway reminds us that it is a mistake to equate decentralization with a lack of control. The internet may be a decentralized system by design, yet the principle of control is nevertheless built into the technical protocols through which the network operates. See Galloway, *Protocol: How Control Exists after Decentralizations* (Cambridge, MA: MIT Press, 2004), 7–8.

11 Other media scholars, such as Amanda Lotz, have explored aggregation as a form of curation. See Lotz, *Portals: A Treatise on Internet-Distributed Television* (Ann Arbor: Michigan Publishing Services, 2017).

12 Andrea Petersen, "What Is a Portal—And Why Are There So Many of Them? Once Gateways to the Web, They Keep Expanding," *Wall Street Journal*, December 10, 1998, B8.

13 Broadcast.com, *US Securities and Exchange Commission Prospectus* (filed July 17, 1998), from SEC EDGAR Database, https://www.sec.gov/Archives/edgar/data/1061236/0000950134-98-006006.txt.

14 David Barboza, "Broadcast.com Soars in Opening Day Frenzy," *New York Times*, July 18, 1998, D1.

2002, and for many, Broadcast.com is today remembered as a cautionary tale, a prime example of the speculative economic boom and bust cycle now known as the dot com bubble. Yet while all in one web portals like Yahoo! and MSN persist to this day, it is AudioNet/Broadcast.com's vertical portal that has proved most popular and culturally significant. For services organized around a particular type of media content, look no further than YouTube, which since its founding in 2005 has become virtually synony mous with user generated online video content.

During the mid to late 1990s, the Broadcast.com vertical portal was among the most trafficked websites in the United States even though the company produced al most no original content. Its business model consisted instead of redistributing exist ing radio content for online audiences.[15] At its peak in 1999, the site featured content from more than four hundred radio stations and fifty television stations, plus game broadcasts for more than 450 college and professional sports teams. It also provided coverage of a wide range of other events, including political speeches, business confer ences, and concerts.[16] In addition to its live simulcasts and webcasts, Broadcast.com offered more than sixty five thousand hours of on demand content with hundreds of audiobooks and nearly twenty five hundred full length music albums in its "CD Jukebox."[17] And all of this media content was available to audiences free of charge. Any broadcaster or other media creator could potentially establish their own website and online stream by the late 1990s, yet Broadcast.com lured in many such content providers, often acquiring streaming rights at little to no cost. Among the reasons that content providers preferred to partner with an aggregator like Broadcast.com was the industry discourse about how people accessed the internet and how businesses could best take advantage of the web. With few proven advertising models in the early years of the web, the industry mostly adopted a crude version of the network era television ad model and tried to attract the largest audience possible. Thus, advertisers sought sites with high traffic, and in particular, sites that were "sticky" and held web surfers' attention for long amounts of time.[18] As Henry Jenkins, Sam Ford, and Joshua Green describe, the "stickiness" online business model "refers to centralizing the audience's presence in a particular online location to generate advertising revenue or sales," which is achieved by "placing material in an easily measured location and assessing how many people view it, how many times it is viewed, and how long visitors view it."[19] This notion of stickiness valued websites that functioned the most like older mass media that maintained attention by their monopoly.

15 AudioNet also distributed some streams from live events, plus music, audiobooks, and eventually TV and video content.

16 Yahoo! Inc., Form 8-K (filed July 10, 1999), from SEC EDGAR Database, http://www.sec.gov/Archives/edgar/containers/fix043/1011006/000104746999028059/0001047469-99-028059-index.htm.

17 Broadcast.com, Form 10-K/A—1998 Annual Report (filed April 27, 1999), from SEC EDGAR Database, http://www.sec.gov/Archives/edgar/data/1061236/0000950134-99-003284.txt.

18 George Anders, "The Race for 'Sticky' Web Sites—Behind the Deal Frenzy, a Quest to Hang On to Restless Clickers," *Wall Street Journal*, February 11, 1999, B1.

19 Henry Jenkins, Sam Ford, and Joshua Green, *Spreadable Media: Creating Value and Meaning in a Networked Culture* (New York: New York University Press, 2013), 4.

The connection between online content aggregation and mass media, especially cable television, was not lost on Broadcast.com's founders, Mark Cuban and Todd Wagner. They frequently compared their vertical portal to a cable service provider. Wagner once proclaimed: "We're just cable on steroids. We're the next step. We're 50,000 channels."[20] In other words, Broadcast.com was not seeking to radically redefine media content or distribution for the internet. The vertical portal may have been a new type of media company for the internet era, yet it was modeled explicitly on broadcasting precedents and imposed those mass media logics on the internet media ecosystem.

The language of web portals may have faded after the early 2000s, yet the portal model and the logic of aggregation remain central within digital media industries and the contemporary online environment. Today we are inclined to call them platforms. Nick Srnicek defines platforms as digital infrastructures that mediate among different groups of users: customers, advertisers, service providers, producers, and suppliers. They are monopolistic firms that provide both the hardware and the software upon which social and economic activity occurs, thereby extracting and controlling immense amounts of data and content.[21] The reach of modern platforms, and particularly the degree to which they are able to extract data, certainly exceeds the portals of fifteen to twenty years ago. Nevertheless, at their core, services like YouTube, Netflix, and Spotify are all content distributors that operate under basically the same intermediation principle of pulling together a vast array of content at a single site from which individuals can then sort and filter the programming according to their particular needs—increasingly, algorithms will presort the content for them on the basis of established preferences. These aggregators have replaced human organized directories as the internet's new intermediaries. The internet has, in many ways, succeeded in opening up the long tail of media content, bringing increased attention—if not economic viability—to so called niche media makers and cultural products. This increased value is the product of new audiences and markets that have opened up through the internet. However, while the mass media production, distribution, and exhibition structures of the twentieth century have been significantly disrupted, it is erroneous to conclude that this is a form of disintermediation wherein producers and audiences are directly exchanging content without an intermediary in the supply chain.[22] This intermediation may occasionally be the case with truly underground filmmakers or indie musicians, as it was in the preinternet era. However, for media content to scale up and reach a mainstream audience, it must still be streamed or sold through an aggregator like Amazon, iTunes, YouTube, Hulu, Netflix, or Spotify.

These new intermediaries and their affiliates maintain as much or even more control over online media distribution and retail than did the broadcast networks or movie theater chains in their respective heydays.[23] Especially in streaming media, the business

---

20 Quoted in Alan Goldstein, "CEO of Dallas-Based Broadcast.com Reflects on Future after IPO," *Dallas Morning News*, August 31, 1998.

21 Nick Srnicek, *Platform Capitalism* (Malden, MA: Polity, 2017), 43–48.

22 Patrick Vonderau, "The Politics of Content Aggregation," *Television & New Media* 16, no. 8 (2015): 720.

23 There are a variety of affiliates involved in the digital media distribution process. One such group are "white label" agent aggregators that bundle content catalogs and negotiate licensing deals and access to branded aggregation platforms like Netflix.

model requires such massive scale for profitability that most industries today can support only a few major aggregators in any given region of the world. For instance, if you are a musician in the United States in 2019 and you want your music to potentially reach a mainstream audience, you have little choice but to work with the aggregators Spotify, Apple Music and iTunes, YouTube and Google Play Music, and Amazon. However, the low royalties paid by streaming music services mean that many artists actually operate at a loss in order to make their songs available through these services.[24] In the radio and podcasting industries, there are no royalties or licensing fees whatsoever; broadcasters and podcasters give away their content free of charge to aggregators like TuneIn, iHeartRadio, Stitcher, and iTunes Podcasts. To turn a profit, producers must grow their audience enough to attract advertising. When advertising is not feasible, as in the case of noncommercial or local radio and podcasting, content producers may still give their programming away for free because they fear they will otherwise lose out on listeners who use only aggregators. There certainly are musicians, radio broadcasters, and podcasters who do not wish to accept the aggregators' contract terms, but their exclusion from these platforms risks rendering them invisible. In this way, content aggregation may actually negatively influence diversity, because the more audiences depend on these platforms as their primary means of consumption, the less likely they are to explore the (unseen and unheard) options that exist beyond the platforms' boundaries. Wendy Hui Kyong Chun has argued that algorithmic recommendation systems are based on assumptions of homophily (the principle that like breeds like), which discourage individuals from stepping outside their comfort zone.[25] In this way, the internet continues to operate as both a vast realm of unrestricted communication and free exchange and a highly consolidated and controlled marketplace.

The internet, and the phenomenon of content aggregation, obscures media content as much as it calls attention to it. There is often a false sense of completeness presented by content aggregators like Spotify or YouTube that offer millions of songs or videos. These platforms offer more media than any one person could consume in a lifetime, and yet they are far from offering all the media that exists, past or present. There are numerous reasons media content may be exempted from an aggregator, many of which come down to economic matters, such as licensing disputes or orphaned content situations, which occurs when no one can find the individuals with the right to license the material. It may be absent because it has been censored for running afoul of a site's "community guidelines." There are also significant cultural biases involved. For instance, access to aggregators like Spotify or YouTube is not globally universal, nor is the content available identical in all regions of the world. Their catalogs differ by region for licensing reasons and also for cultural reasons. Perhaps unsurprisingly, in the United States there is a predilection for American centric media content, and

24   Patryk Galuszka, "Music Aggregators and Intermediation of the Digital Music Market," *International Journal of Communication* 9, no. 1 (2015): 268.

25   Wendy Hui Kyong Chun, *Updating to Remain the Same: Habitual New Media* (Cambridge, MA: MIT Press, 2016), 15.

foreign content is often excluded outright.[26] Language is turned into a significant sorting mechanism, as content aggregation platforms in English speaking countries limit the access and discovery of non English language content.

Since the earliest days of the internet and, especially, the World Wide Web, aggregation has played a significant role in ideas about media distribution, access, and consumption in our networked culture. Socially and politically, these ideas about massive collections of heterogeneous media content are central to optimistic claims about the medium's democratizing effects and the internet's ability to increase democratic choice and participation. Yet the consolidation of content through aggregators raises concerns about concentrated power and control.

---

26 If not excluded outright, foreign content is minimized via algorithms that push users toward trending content, or it may be relegated to inferior-quality streams, and so on.