Digital hoarders: “Our terabytes are put to use for the betterment of mankind”

reader comments

92 with 67 posters participating

Share this story

[embedded content]
Think we prefer the album version, but OK, sure *Top of the Pops*
Today perhaps more than ever, data is ephemeral. Despite Stephen Hawking’s late-in-life revelation that information can never *truly* be destroyed, it can absolutely disappear from public access without leaving a trace.

It’s not just analogue data, either. Just as books go out of print, websites can drop offline, taking with them the wealth of knowledge, opinions, and facts they contain. (You won’t find the complete herb archives of old Deadspin on that site, for instance.) And in an era where updates to stories or songs or short-form videos happen with the ease of a click, edits happen and often leave no indication of what came before. There is an entire generation of adults who are unaware that a certain firefight in the Mos Eisley Cantina was a cold-blooded murder, for instance.

So on any given day, 19-year-old Peter Hanrahan now spends his evenings binging on chart-topping radio shows from the 1960s. A student from the North of England, he recently started collecting episodes of *Top of The Pops*—a British chart music show which ran between 1964 and 2006—after seeing the 2019 Tarantino flick, O*nce Upon a Time In Hollywood.*

“I was searching for TOTP episodes as I found that there was a severe lack of them available on YouTube, the BBC iPlayer, or any other radio shows,” he tells Ars. “But I wanted to experience what it would have been like back then and searching because of how atmospheric the radio was in *Once Upon A Time in Hollywood*. It’s been another way to discover music from that era.”

If Hanrahan merely wanted to experience more ‘60s British chart-toppers, of course, he could’ve simply run to Spotify. But he wants the experience of TV as it was recorded back in the day—including live studio audiences, lip sync controversies, and alleged sex offenders.

Naturally, YouTube does have many old episodes, but the BBC has tried taking down ones featuring Jimmy Savile or Gary Glitter, for instance. Today it’s far from a complete TOTP library with only a fraction of the episodes Hanrahan is looking for accessible on the platform. YouTube is also quick to respond to takedown notices, and episodes which are currently there one day can disappear the next.

His next stop is archive.org, the venerable non-profit library which boasts a tremendous 411 billion archived web pages, 23 million books, 5.5 million movies and a variety of other data. Often they will have what Hanrahan needs, but if not, his next stop is an obscure corner of reddit, where it is just possible that someone, somewhere, will have a copy saved.

It’s taken Hanrahan a long time to find and obtain them, but his work, trawling the edges of the Internet and connecting with real people, is finally paying off. In his first year as a self-confessed hoarder, Hanrahan had collected more than than a terabyte of data.

This impermanence of information, of course, goes far beyond old British radio. And luckily for future generations, the itch to seek it out, collect it, and store it goes beyond

Hanrahan, too. It’s a sentiment currently driving thousands of individuals to band together online in the communal pursuit of archiving old media of all sorts. This ain’t the grant-and-partnerships-funded well-coordinated operation of the Internet Archive; it’s the individual-obsession-driven r/Datahoarder.

There’s a subreddit for everything

In 2020, the r/Datahoarder community on reddit is almost 200,000 members strong, with around 1,000 or so idling or posting in the subreddit at any time. The communal purpose here is exactly what it sounds like: these amateur archivists set out to collect and capture data and to preserve it for record, reference, and future reading. Often, the goal is to retain this information both online and off, through physical media or terabytes of personal harddrives and storage. In a way, you can think of r/Datahoarder like thousands of haphazard individual Internet Archives—though each member tends to have a few specific niche areas of focus.

On r/Datahoarder, you’ll find people storing data on everything from YouTube videos to game install discs. One person was even planning to copy all Australia-based websites even as the country burned in the worst wildfires in history. The post was deleted after it was pointed out that the physical servers for Australian websites are located outside the country. They’re safe for now—phew.

Some users archive every website they visit or service they use, and the gamut of media includes virtually everything: movies, music, and porn are all popular.

And for future historians, every tweet, every livestream, every TV and news show of the recent and ongoing Hong Kong democracy movement has been squirrelled away by a few dedicated users. Already it’s proving useful to at least one academic who visited r/DataHoarder seeking research material for their Sociology master’s thesis on the Hong Kong protests.

Any hardware is welcome. While many users boast huge storage racks of expensive equipment, even humble Raspberry Pis are routinely kitted out with oversized drives and employed as real-time reddit-scrapers. That embarrassing 3am post about how you really need to get back with your ex? You may have deleted it within seconds of posting, but it's almost guaranteed that there are multiple copies in private archives—available to your ex on request.

1990's era mass storage devices such as the Iomega Zip Drive occasionally float to the surface of the sub, as their owners rediscover them from a cupboard under the stairs, prompting discussion on drivers, recovery methods, file formats, and readability.

The desire to save information for posterity seems to be almost universal, but manifests in different ways according to each hoarder's own interest. Scroll through the boards and you'll find archived websites offering customization for Windows 98 machines and novelty cursors. You'll find users on a mission to preserve the entire Internet of a single country at a given point in time. You'll find users whose particular obsession is satellite weather forecasts for Japan, or silent movies.

As you might guess based on a collection of highly motivated and obsessed tech users, r/Datahoarder started first as a single IRC chat channel on freenode. Eventually, the community transitioned to the still-in-occasional-use r/datahoarders, with r/datahoarder being brought into existence four years ago. There is also a separate exchange subreddit, r/DHExchange, where members attempt to fill gaps in their collections.

Discussion these days is typically highly technical, largely revolving around efficient means of storing or hoarding vast quantities of data gleaned from online and elsewhere. Users want to get advice on hard drive arrays running into the

hundreds of terabytes, mass storage options in the cloud, and the astonishing costs associated with archiving otherwise forgotten older media like broadcasts, music, journals, and webpages.

Hanrahan didn’t get involved out of his love of the 1960s musical zeitgeist—old British music acts are only the latest archival effort he’s undertaking. In real life, Hanrahan has 12 drawers of color-coordinated Lego bricks he uses frequently and an extensive vinyl collection, which includes everything from the original *The Good, The Bad, and The Ugly* soundtrack to music from *Red Dead Redemption II*. Perhaps unsurprisingly, he also maintains a large digital games library.

“It started out as me compiling together stuff that I think is relatively hard to find, and just some cool stuff I find, like old commercials and TV intros like ABC’s,” he said.

As a small and whimsical fish in the data hoarding pool, Hanrahan’s storage isn’t extensive but is still considerably more than what most users would have on their home systems. His storage capacity is 6 TB, with 3 TB given over to backups. He spends an additional £100 (roughly $130) on two 1TB drives each time he starts to run out of space. He even keeps additional drives containing his most valued data at another family member’s house and updates his hoard yearly.

Enlarge / No, not that kind of hoarder. Save that for AMC.
Kurt Wittman/Education Images/Universal Images Group via Getty Images

A brief history of archiving impulses

The urge to store rare or useful recordings and information has been going on for as long as humans have had the means at their disposal. The first archives of written material started appearing at around 3500 BC—not long after the invention of writing, and the Great Library of Alexandria was founded with the aim of acquiring and hoarding the best and most authoritative copies of every piece of work ever produced, employing scribes to hand copy onto the finest parchment available—the ancient equivalent of 8K UltraHD blu ray rips.

It wasn’t until the 1970, with the phenomenal success of the compact cassette tape that amateur archiving of popular live media became possible. Teenagers in their bedrooms would

record live radio shows as they aired with the latest pop songs from pirate radio stations. By 1974, Billboard magazine reported that over 40 percent of all age groups recorded live shows from the radio, with a corresponding drop in the number of prerecorded tapes being purchased. Home taping is killing the music industry? This is where it started. Tapes were recorded and recorded again, before being condemned to disposal or a purgatory of eternal storage in a slowly yellowing plastic case, or at the back of a kitchen drawer.

The advent of Betamax and VHS soon gave hoarders a new tool. Live and pre-recorded TV shows and movies became available to watch on demand from the users' own personal libraries. As with cassette tapes, most recorded shows were later recorded over to make room for the next episode of *The Bob Newhart Show* or *All In The Family*. What most people had in mind was not a permanent archive—it was the convenience of being able to watch or listen to the latest installment of a favorite soap when it suited them.

But as VCRs gave way to DVD players, then to DVDRs, TiVo boxes, and eventually the streaming landscape we know and love today, VHS tapes suffered the same fate as cassettes. Broadcast TV, like radio, has largely been lost to the mists of time unless the creators and rights holders put in the effort to create and securely store backups.

For instance, *Doctor Who* is one one British television's most successful exports, and at its peak popularity in 1982, the show was being watched by a global audience of 98 million people. Today, the fandom is obsessive—poring over the tiniest plot details, stockpiling episodes, and arguing over which of the Doctor's 13 incarnations was the greatest.

But between its initial broadcast in 1967 and 1978, the BBC routinely deleted its programming after it had been broadcast in the belief that there was no practical value to keeping copies. Nine years of beloved *Doctor Who* episodes are missing.

Some clips survive and occasionally, a full episode will turn up, courtesy of a foreign network which found the original two inch tape in a box down the side of the couch, but most of *Doctor Who*‘s earliest broadcasts are gone for good.

Listing image by MARTIN BUREAU/AFP via Getty Images

https://arstechnica.com/?p=1667508