

PRESERVATION

Saving for the future

A growing proportion of global culture exists only online, presenting a challenge for those tasked with maintaining the historical record.

BY SEDEER EL-SHOWK

In August 2017, Hurricane Harvey swept through the Caribbean before making landfall in southern Texas and Louisiana. It led to more than 100 deaths and caused an estimated US\$125 billion of damage in the United States alone. As recovery efforts began, researchers and archivists from Rice University in Houston, Texas, together with the Houston Public Library, Harris County Public Library in Houston, and the University of Houston Libraries, set out to create what they describe as a digital memory bank of the storm.

Numerous photos, videos, audio clips and stories from affected communities had already been posted on social media, but there were no guarantees that they would remain available for posterity. “You don’t want all that stuff to get lost and never preserved or archived in a way that future generations can access and learn from,” said Caleb McDaniel, a historian at Rice University who is part of the project.

The Harvey Memories Project aims to process and catalogue this material in a permanent archive, preserving the communities’ experience of the hurricane for historians and other researchers. The archive launched in July and already houses hundreds of records, but the team hopes to save tens of thousands.

The Harvey Memories Project highlights a growing problem: much of our cultural experience is now mediated by ephemeral technologies. Hundreds of millions of photos are uploaded to social media every day, and an ever-growing portion of our cultural output, from memes and cat pictures to tweets, podcasts and educational videos, exists only online. Archiving these digital materials poses a host of technical, legal and social challenges, many of which are exacerbated by the fact that



much of the material is in the hands of private corporations such as Facebook and Google. These challenges raise important questions for anyone concerned with preserving our cultural heritage.

“We must start talking about what values and principles we want to guide the curation of historical records: generational justice, scientific, religious, commercial,” says Carl Öhman, a researcher at the Oxford Internet Institute’s Digital Ethics Lab in Oxford, UK.

PLAYING SAFE

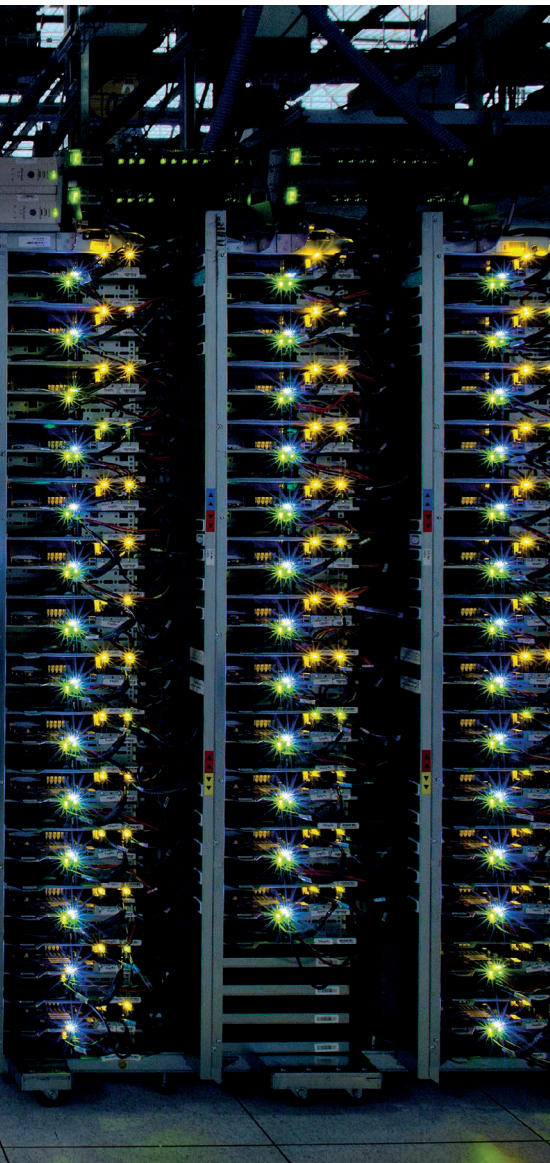
A cross-disciplinary array of experts is working out how to address these questions. Earlier this month, the International Internet Preservation Consortium held its annual Web Archiving Conference in Wellington, New Zealand. The meeting brought people from a variety of disciplines together to discuss the social and technical obstacles to preserving the world’s online heritage. Talking points included the development of new tools for collecting online media, and the difficulties encountered when

dealing with transnational platforms.

One of the technical challenges facing archivists is choosing a storage medium that will stand the test of time. Just as floppy disks disappeared and optical disks are becoming less common, modern storage media such as memory cards and USB sticks are likely to be supplanted by newer technologies. Disks and drives also eventually wear out because of physical and chemical degradation. To safeguard access to stored information in the face of decay or technological obsolescence, archivists regularly transfer data to new media. But errors can creep in with each transfer. To spot them, archivists create a ‘digital fingerprint’, known as a hash, before copying the original file. This string of letters and numbers is unique to that file and can therefore be used to verify that any copies are identical to the original. If the file is changed during copying, its hash will no longer match, alerting archivists to the need to try again.

Corporations such as Google store enormous amounts of data.

GOOGLE CLOUD



In some cases, digital objects need to be deliberately modified to preserve them, for example by removing noise from an audio recording. But even then, archivists also keep the original master whenever possible.

“One of the reasons for maintaining the digital original is that in 20 or 30 years’ time there might be a mechanism where we could actually go back to the original and use it in a manner that we can’t now,” says Steve Knight, head of the digital-preservation team at the National Library of New Zealand in Wellington. Archivists use various equipment and techniques to preserve the integrity of the original, such as write-blockers, which prevent a computer from writing to a connected hard disk.

These hardware problems are compounded by rapid changes in software and file formats. It is possible to replicate the outdated software required to view a particular type of file on modern equipment, but this can involve significant digital sleuthing if little is known about the original file format or software. In 2013, when enthusiasts at the Carnegie Mellon

University Computer Club in Pittsburgh, Pennsylvania, recovered a cache of image files from Amiga floppy disks that belonged to pop artist Andy Warhol, for example, they spent months reverse-engineering the software needed to view the images. Their effort was rewarded with a digital reproduction of Warhol’s famous soup cans and other digital experiments. Once recovered, such files can be converted into a modern or standardized format, although this might result in the loss of properties or information embedded in the original, such as metadata recording the location at which photographs were taken.

The final technical hurdle is ensuring that the provenance of the data is recorded for use by future scholars. Digital objects are more vulnerable to tampering than traditional artefacts, but the verification and preservation tools used during copying enable archivists to prevent or detect malicious manipulation. “The integrity and authenticity of the digital object is at the root of the digital preservation endeavour,” says Knight.

ACCESS ALL AREAS

If national libraries are to serve as the memory of a nation and provide what Knight calls “a communication line with the future”, there must be a mechanism to let them access the material. For printed documents such as books and periodicals, many countries have laws that require publishers to provide copies to their national libraries.

In 2003, New Zealand became one of the first countries to extend this principle of legal deposit to digital objects. This gave the National Library the right to archive websites based in New Zealand and other digital material created in the country, and allowed it to bypass copy protection to preserve the data, provided that copyrighted data are not made accessible without permission. Many digital data are beyond the scope of such laws, however. In particular, much of the information that archivists want to preserve is in the hands of large international corporations that might have little interest in cooperating with libraries.

For example, much of the music produced by New Zealanders is hosted on online platforms such as Bandcamp, says Knight, and these have little incentive to deposit their audio files with the National Library. The transnational nature of social media and other online services means that “a lot of those activities are effectively happening offshore”, he says. “This brings up a whole range of not just legal issues, but also social and cultural questions around how national institutions build and protect the digital collections chronicling the history of their countries.”

Collaborative efforts between national libraries could help them reach beyond national boundaries. But these approaches can be stymied by legal differences between nations, ranging from what material is covered by legal

deposit, to laws regarding libel, obscenity or blasphemy.

Libraries can also face problems with legal deposit when people publish images and text online while travelling abroad. Legal-deposit laws vary between countries and do not always authorize the collection of digital material published by its nationals outside the country. As a result, national libraries might sometimes need to determine whether a digital item was published from within the country or abroad, making the collection process inordinately complicated. Knight suggests that archivists should proceed boldly, with a mind-set of seeking forgiveness later rather than permission in advance.

The sheer quantity of digital information published by companies and individuals also makes collection difficult. In 2010, the US

“You don’t want all that stuff to get lost and never preserved or archived.”

Library of Congress reached an agreement with Twitter that enabled it to archive every tweet since the company’s inception in 2006. The library announced in 2013 that it had collected all the tweets

from 2006 to 2010 and established a process for managing the continuous incoming stream, which had grown to roughly half-a-billion tweets per day. But the library recently changed its collection policy: from the start of 2018, it started archiving tweets selectively, as a result of the continued growth in the quantity of posts and the number of images and videos being shared. Until a way can be found to provide access cost-effectively, the contents of its now more limited Twitter archive will remain under embargo.

The selection of tweets for the archive will follow the library’s general collection guidelines, which focus on preserving material related to events of national interest. However, the need to be selective raises important questions about which materials are preserved in a nation’s memory. Digital technologies should make it easier for smaller or marginalized communities to be heard, but this diversity is still not always captured. Libraries are not entirely neutral repositories of knowledge; intentionally or not, the choices made about what to preserve reflect society’s inequalities and biases (see page S147).

PERSONAL PROBLEMS

Whereas libraries are forced to make difficult choices, social-media companies have the capacity to store vast quantities of our personal digital information in their data centres. These enormous private archives are not managed with the aim of preservation that guides public libraries, but they nevertheless have incentives to retain users’ data — even when the user has died.

The accounts of deceased people are commercially valuable as long as they continue to



The Harvey Memories Project is preserving the public's images, videos and stories of Hurricane Harvey.

generate interest and activity from friends and family. Facebook and Google have policies that enable users to determine how their account should be managed after their death. These memorials might help to preserve material that is not kept by libraries, but their longevity is dependent on their commercial value.

The management of digital remains creates a new set of legal questions. "These in-service solutions are partial and sometimes problematic," says Edina Harbinja, a senior lecturer in media and privacy law at Aston University in Birmingham, UK. For example, they might clash with a will or inheritance laws. "A friend can be a beneficiary for Google or Facebook services, but they would not be heirs and next-of-kin who would inherit copyright on one's assets," she explains, leading to confusion if the account contains copyrighted material.

The laws governing privacy and succession also differ between countries, and this could further complicate the interpretation and implementation of these policies. Harbinja sees them as a start towards a more comprehensive system of 'social-media wills' as laws regarding digital remains develop and, ideally, become harmonized across nations.

Despite these efforts, the fate of our digital remains can still pose problems. In 2012, a 15-year-old German girl was killed by an underground train. Her parents asked for full access to her Facebook account — not simply a memorial site — in the hope that it would hold clues about whether her death was a suicide, perhaps resulting from online bullying. An initial court ruling in 2015 granted them access, but the decision was overturned on appeal in 2017. This debate centres on whether the girl's contract with Facebook can be inherited by her parents in the same way as letters or a diary, and whether this would violate privacy laws.

In July 2018, Germany's highest court ruled in favour of the parents, determining that

social-media accounts should be passed on to heirs in the same way as books and letters. Harbinja disagrees with the decision, believing that the court overlooked some fundamental ethical questions. She argues that a contract with Facebook is purely personal and that the inherent right to privacy should extend beyond an individual's death, regardless of the circumstances. Moreover, granting heirs access to an account would give them the ability to view material shared privately by contacts of the deceased person, violating the right to privacy. "Online representations of self and identity are much more complex than one's

letters and pictures," Harbinja told Australian radio, advocating a nuanced approach, rather than a one-size-fits-all solution.

Alongside the legal questions, the commercial management of digital remains means that their use will be driven by profit incentives. This might lead to their commodification, or the exploitation of the grief of the bereaved (see 'Digital immortality'). Öhman and Luciano Floridi, also at the Digital Ethics Lab in Oxford, advocate that digital remains should be treated as an extension of physical remains. "Data is not merely something we own, like a car, but something we are, like an arm," says Öhman. "When someone intrudes on our privacy, we don't lose anything that we own, but we may lose control over who we are, our dignity. It follows that since our privacy can be violated without our knowledge, it can also be violated when we are dead."

As our lives become increasingly enmeshed in digital communications, questions about what we choose to preserve and how we manage those materials will play an ever-bigger part in the formation of our cultural heritage. Archiving social-media posts might seem trivial in the wake of a hurricane, but it is driven by the same motivation that made libraries collect the newspaper clippings and first-hand accounts that expand our knowledge of disasters a century ago. New technology has brought fresh challenges, but few dispute the need to preserve our otherwise ephemeral recordings. ■

Sedeer el-Showk is a science writer based in Finland and Morocco.

SOCIAL MEDIA

Digital immortality

When you die, the way your digital remains are handled inevitably raises questions about dignity and exploitation. Conventional social-media companies go no further than turning a profile into a memorial, but 'digital-afterlife' start-ups such as Eternime and Eter9 offer a more ambitious alternative. Given access to your social-media accounts, their algorithms will analyse your images, links, posts and interactions to build a 'virtual you' — a digital representation of your online persona that will eventually interact with your loved ones. Neither of these services are live yet, but tens of thousands of people have signed up, highlighting the allure of digital immortality.

Ethicists warn of moral quandaries surrounding these digital recreations. "If firms compete in making the dead 'consumable', our memory of the dead will be guided only by the principle of profit, and not principles of justice, historical value,

sentimental value and so forth, unless such principles happen to align with what consumers want," says Carl Öhman of the Digital Ethics Lab in Oxford, UK.

The chatbots used by such sites might also gradually diverge from the original persona. Financial incentives could push companies to calibrate chatbots towards commercial goals that might be at odds with an honest depiction of the deceased. For example, because engagement is an important commercial metric on social media, the bots might be more extroverted or chatty than the original person.

These concerns lead Öhman to suggest that digital-afterlife companies should have to ensure that consumers know how their data will be displayed post-mortem, that users will not be depicted radically differently from the original bot, and that people can upload only their own data, not data to create a representation of friends or relatives. **S.S.**