

# America archives its billions of tweets

January 22 2013, by Anne Renaut

---



Photo taken on October 8, 2012 shows the main reading room at the Library of Congress in Washington, DC. The Library of Congress, repository of the world's largest collection of books, has set for itself the enormous task of archiving Americans' billions of tweets.

The Library of Congress, repository of the world's largest collection of books, has set for itself the enormous task of archiving something less weighty and far more ephemeral—Americans' billions of tweets.

The venerable US institution is assembling all of the 400 million tweets sent by Americans each day, in the belief that each of the mini-messages reflect a small but important part of the national narrative.

"An element of our mission at the Library of Congress is to collect the story of America, and to acquire collections that will have research value," according to Gayle Osterberg, director of communications at the library.

The Library of Congress, located off the National Mall in Washington, houses millions of hard copy books and historic documents, and its online archives amass millions of additional works produced by Americans for more than two centuries.

Now it wants to be keeper of the nation's brief Internet messages as well: [Twitter](#) in April 2010 inked a deal with the Library, giving it access to tweets dating back to the company's inception in 2006.

Collecting the 140-character micro-missives, said Osterberg, is in keeping with the library's main goal "to collect the story of America and to acquire collections that will have research value."

One major challenge to the Library, however, is storing the messages from the popular social messaging site, which now number 170 billion. Twitter last month said the number of active users on the messaging platform has topped 200 million, most of whom are in the United States.

Tweets that have been deleted or that are locked will not be among those gathered by the Library of Congress.



The Library of Congress is assembling all of the 400 million tweets sent by Americans each day, in the belief that each of the mini-messages reflect a small but important part of the national narrative.

Among the messages to be preserved for posterity are the first-ever tweets sent by one of the company's founders, Jack Dorsey.

Also saved for all time is a famous tweet sent by President [Barack Obama](#) after his historic November 2008 victory to claim the White House in his first term.

"We just made history. All of this happened because you gave your time, talent and passion. All of this happened because of you. Thanks," read the micro-message from the famously tech-savvy US president.

Unlike traditional bound books or even digital web pages, the real challenge of preserving tweets is keeping up with their number, which

has continued to grow almost exponentially.

There were 140 million tweets sent each day in February 2011, but more than three times as many—about a half billion—by October 2012.

The Library of Congress's tweets are being stored by Gnip, Inc., a social media aggregation company headquartered in Boulder, Colorado, which has put more than 133,000 gigabytes of storage space available.

Gnip says it is a particular challenge to gather tweets during "peak" times, such as news event watched the world over like the Japanese tsunami in March 2011, which generated many thousand tweets per second.

It has proven to be a Herculean challenge for Gnip to make tweets accessible to all those who wish to view them.

So far it has been unable to meet the demands of researchers worldwide who hope to access the archive. Even a search among the first four years of [tweets](#), from 2006 to 2010, could take about 24 hours.

"It is clear that technology to allow for scholarship access to large data sets is lagging behind technology for creating and distributing such data," said a recent white paper published by the [Library of Congress](#).

"This is an inadequate situation," the Library concluded, calling the massive archiving project "prohibitively costly."

And yet Lee Humphreys, a professor of communication at Cornell University in New York, said that the brief online messages can reveal volumes "about the culture where they were produced."

(c) 2013 AFP

Citation: America archives its billions of tweets (2013, January 22) retrieved 19 June 2024 from <https://phys.org/news/2013-01-america-archives-billions-tweets.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.