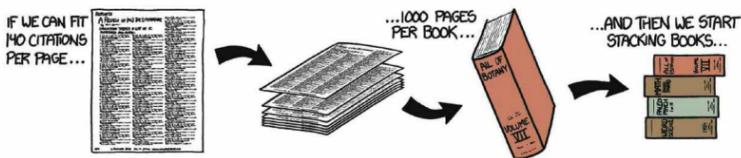


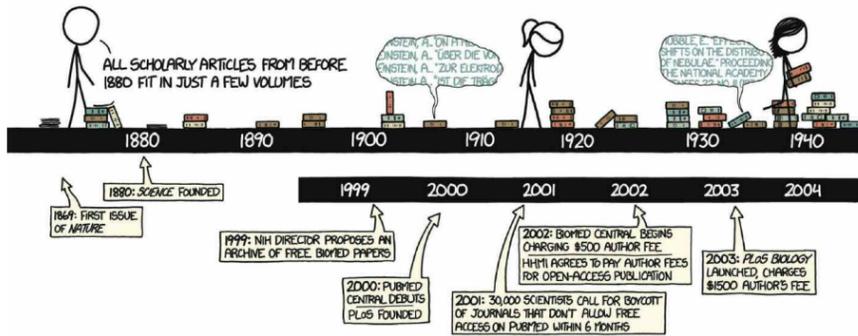
## HOW MUCH SCIENCE IS THERE?

SCIENTIFIC PUBLISHING HAS BEEN ACCELERATING—A NEW PAPER IS NOW PUBLISHED ROUGHLY EVERY 20 SECONDS. LET'S IMAGINE A BIBLIOGRAPHY LISTING EVERY SCHOLARLY PAPER EVER WRITTEN. HOW LONG WOULD IT BE?



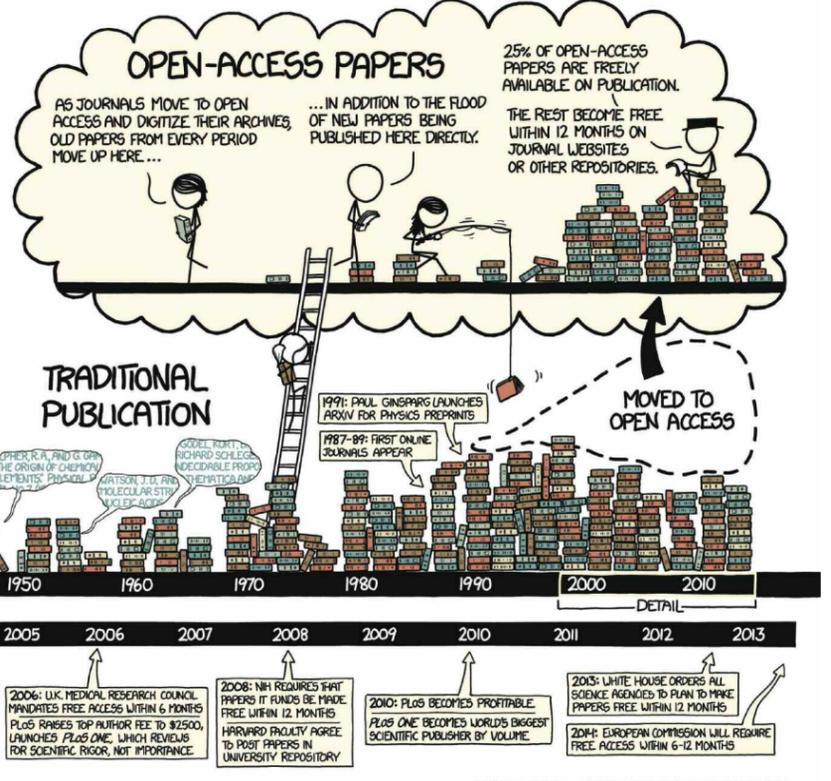
A LIST OF PAPERS PUBLISHED IN 1880 WOULD FILL 100 PAGES.  
BY 1920, THE LIST WOULD BE GROWING BY 500 PAGES PER YEAR.  
THE 1975 SECTION WOULD FILL FOUR HUGE VOLUMES.  
TODAY, WE'RE UP TO 15 VOLUMES PER YEAR—A PAGE EVERY 45 MINUTES.

...THIS IS WHAT THE FULL LIST WOULD LOOK LIKE:



## HOW OPEN IS IT?

SINCE THE ADVENT OF THE WEB, MUCH OF SCIENTIFIC PUBLISHING HAS BEEN MOVING TO OPEN ACCESS. ACCORDING TO SCIENCE-METRIX, OPEN ACCESS REACHED A "TIPPING POINT" AROUND 2011: MORE THAN 50% OF NEW RESEARCH IS NOW MADE AVAILABLE FREE ONLINE.



Drawing xkcd.com

### The data archive of the Sudan census of 1973

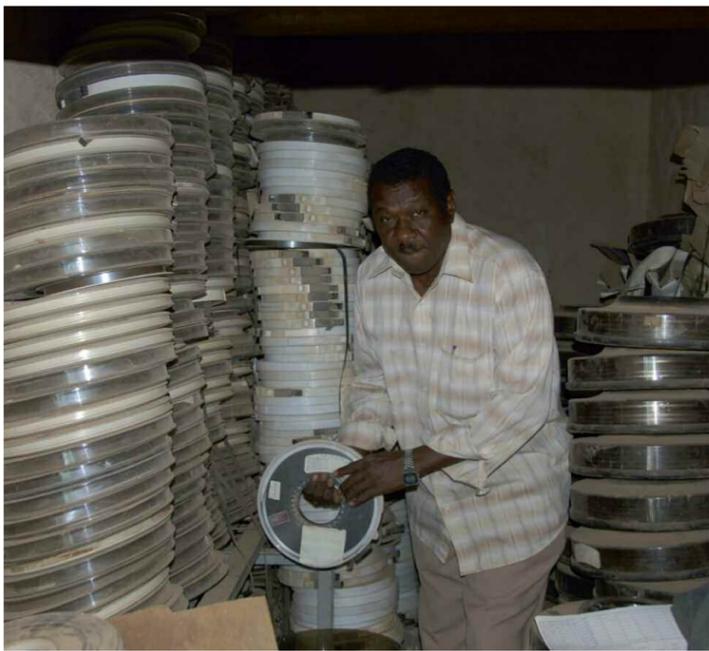


photo McCaa, Research Professor, Ambassador, IPUMS

# Enhanced data journal: next generation science

**DANS is about to launch its own digital data journal. Leen Breure explains the initiative.**

A digital data journal will contain data papers and reviews of datasets. A data paper is a scientific article describing data and their production against the background of the research project in which they were created. Some form of peer reviewing is usually part of the publication process. One essential property of such a paper is that it can be formally cited and provides credits to the data creators, thus stimulating researchers to make their data more easily available to others.

### Enhanced publication

The DANS data journal is an enhanced publication in more than one respect. The text is enhanced with direct links to datasets stored in the archived repository. Additionally, the journal is enriched with features that contribute to greater usability of the content in terms of overview and navigation by adding background information and various forms of visualization. Where possible, data can be previewed and explored online

rather than through time-consuming downloads and offline applications. In short, an enhanced data paper provides an integrated view of data in their research context.

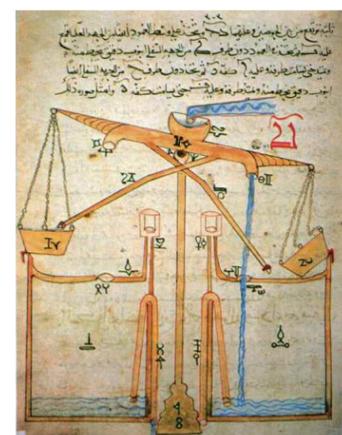
### Mixed blessing of print

This enhancement could easily be considered a novelty that is essentially a luxury. After all, what really matters is some form of access to the dataset itself. However, looking back at the history of scientific publishing we see that integration of scholarly discourse and data – or evidence – has a long tradition. Remarkably, it has been the reproduction method

which has often disturbed this integration. Medieval scientific manuscripts used to have diagrams and pictures, surrounded by notes and directly linked to the textual content. Printing created a problem: wood cuts are coarse and Renaissance anatomists and microscopists opted for 'high resolution' in the form of etching and engraving. But this so-called intaglio printing could not be directly combined with the relief printing used for text. The illustrations were therefore printed on plates at the end (or in the middle) of the publication, separated from the body text. Something similar happened to data.

In early reports of scientific experiments text is mixed with large tables and data lists (statistical graphics are a late 18th-century invention). At that time historians used to create disproportionate footnote sections with lengthy source citations. Massive computerization created not only much more data, but also posed the problem of relating printed and digital information. Finally, by fully embracing digital publishing we are now able to do justice to this natural coherence of text and data and take the next step in an old tradition.

Leen.breure@dans.knaw.nl



Picture of Al-jazari water device with Arabic handwriting credit commons.wikimedia.org/wiki/File:Al-jazari\_water\_device.jpg#mediaviewer

### Sharing data: good for science, good for you >>

8. All your data, preserved securely and sustainably. It adds extra value to your research, to your significance as a scientist. You even increase your chances of citations.
9. Martijn Kleppe: "Anyone who uses my data, should refer to me. That's another enjoyable reason to share my data."
10. Marion Wittenberg: "Also for verification. Researchers should make their data available so that others can check their modus operandi."
11. The repository had to meet requirements for certification and one of them is DSA, Data Seal of Approval.
12. Manfred te Grotenhuis: "You always want more. Ideally, all collected data should be available for further analysis. So, I would really like everybody to make their data available."