20 years Laborjournal

Digital Stone Age

From Björn Brembs, Regensburg

Blocked literature, as it criticizes is d by the "Open Access" movement, is just the tip of the iceberg. Rather, there is a lack ~~of science~~ in the sciences in general ~~to~~ regarding the digital infrastructure that document their "output" in an efficient manner and would make it usable. A halt, which could be dangerous soon.

"Open Access" (OA) is currently on everyone's lips. On the one hand ~~sprout~~ are risen with more and more subscription-independent journals ~~in the leaves of the forest now~~ more than 30,000 peer-reviewed journals (the so-called "golden" road to OA). On the other hand demanding more research funding worldwide that grant recipients make their publications in institutional repositories available (the "green" road to OA), if they are not gold-OA be published. In Germany, the copyright was exactly that extended to the secondary publication rights. In Baden-Württemberg the legislature this possibility now also planning to expand to University employees. But the German Association of University Professors, otherwise reliable representation of interests of the scientists is reflected in a spectacular change of position on the side of the corporations and is preparing a lawsuit against the law in Baden-Württemberg and thus against a development of our digital infrastructure.

One is tempted to exclaim with Shakespeare: much ado about nothing! Because although the accessibility of science has by no means improved over the last 15 years, the OA movement - quite the contrary! -, Then but the entire digital infrastructure at public research institutions in the last 20 years developed more almost. This is now such a variety of shortcomings has accumulated that has completely pushed the problem of literary access in daily research work, depending on the field partly in the background. The stalemate has become a serious domino effects, which not only endanger the preservation of our research results, but also the continued existence of public research as a whole.

Depending on the field fall the fruits of scientific research at one or more of the following three categories: scientific source code, digital data or text-based content.

The lack of accessibility of the literature is only one of many problems of our text-based results:

- Depending on the department, four or more engines must be used to ensure adequate coverage of the literature (in my case the neurobiology: Google Scholar , PubMed , Scopus and Web of Science).
- Although hyperlinks were presented in 1968 by Stanford University for the first time, they have kept still no way into our literature, even after almost 50 years. Or have you ever get an accurate description of the experimental procedures, when you click " the experiments were Performed as Described added anonymously clicked "?
- We will send the journals still images with curves, graphs and diagrams, even if the publishers but would only have our data and some commands to create the diagrams. By the way that would allow the evaluators and later readers to take other aspects of the data in inspection, as selected by the authors.
- We have to rewrite our articles in part radically still after each rejection, because each journal our texts would love to have in a different form.
- Only now a few journals begin with a technology that every student is already built in the 1990s in his websites: Counter for the number of accesses. After all, this technology is only about 20 years old, and not nearly 50 as hyperlinks.
- The publishers actively block and pure private gain access to modern research methods such as content mining.
- Although online retailers such as Amazon offer for over a decade, derivatives, related to products already purchased, there is a comparable technology for scientific articles only in a very limited sense. There is not a single digital tool that makes it easier for a scientist, although these technologies are already being used for many years by scientists in detail in non-scientific fields of the newly published literature individually and from user behavior learning to filter, sort, and to discover .
- There is no scientific assessment options. The much-quoted "impact factor" is about as scientific as dowsing or pendulums. The data of the past 20 years even suggest that dice is appropriate to find a good article in a selection, as these negotiable, non-reproducible and mathematically calculated number wrong.
- We have no ways to apply the new social media technologies on our literature. While slowly disambiguation of authors ORCID is developed - but before this implementation is established on a broad front, many years will pass.

These ten examples are of course only a small selection from the ever-growing number of functionalities that we expect these days of digital objects for granted, however, do not find yourself in our literature. It could be indeed argue that the public availability of so obviously antiquated Material: Since the fact that publishers now actually do with their blockades the exact opposite of what her

carved in stone, ablichteten with the digital cameras in our mobile phones and then put the pictures on the net.

And as if it were not bad enough to the literature, the situation in scientific source or scientific data does not look much better. Only about 25% of the scientific data are available at all, and only a small part of it again in public databases. The remaining 75% will die with their researchers, if they ever survive so long. The few percent that are available in public databases (in biomedical research, especially sequence data), are under the constant threat of financial collapse of these databases. Most of the well 1,400 biomedically relevant databases are operated with project funds from funding to funding. Self-funded by major governments databases are not immune to financial problems, such as the "Government Shutdown" of the U.S. government all too clear recently made. A few days longer political chaos, and a large part of the global biomedical research could not have come to a standstill. In short, for more than 30 years, we generate more and more and more and more scientific fields digital data of inestimable value, but there is no sustainable infrastructure that preserves this data global, long-term disaster-proof. It's as if the floor plans and blueprints actively destroyed with the completion of a house building.

For scientific source code the first tentative steps were taken towards a digital infrastructure until this year. The CERN in Geneva, FigShare (Macmillan / Nature Publishing Group) and the Mozilla Science Lab have coordinated with GitHub and CrossRef and can now Digital Object Identifiers (DOIs) assigned for the source code and make it so quotable. Apart from these initiatives as well as some projects scattered around the world (for example, the DFG-funded SciForge) there is only subject-specific point solutions, in which are to exchange directly colleagues scripts and code and make it accessible.

This problem is of course of similar scope as the problems of literature or data. Scientific code not only serves our experiments from Drosophila flight simulator on the fMRI brain scanner for the LHC, but also evaluates the data collected from leaves and computer models of the atom to the brain or through the air. There is currently no institutional way to preserve this work sustainable, standardized, accessible, and thus to make nachnutzbar. Especially in projects of public interest, such as climate models, which is a fatal and untenable situation.

The lack of functionality (and in some cases non-existent) but our digital infrastructure is only one of three main aspects, among which we have to consider and develop our infrastructure.

In addition to the functionality of the incentive structure is an essential component of our infrastructure. Due to the enormous overproduction of university graduates, compared with academic research positions, creates a morbid competition. Today it is no longer sufficient to operate good research and to collect reliable data. No, you have to accommodate these data are also available as high up in a journal hierarchy that any empirical basis in fact. This means that it is more important where you published than what you published. If you then also pulls the empirical data into account, according to which the methodological quality of the work with the height in the journal hierarchy decreases rather than increases, it is not surprising that with the reputation of the journal not only the proportion of later retracted articles increases, but the number of three test fraudsters who publish there. (See also Alexander Lerchl essay in this issue: " How does the system favors research fraud ")

Sensationalism and low quality claims in the top journals, coupled with the despair of having to accommodate a publication in one of the decisive journals, is the perfect recipe to accommodate the least reliable science in the most respected journals. This system, we not only Jan-Hendrik Schön, Woo-Suk Whang or Diderik owe stack, but also the cover story, "How Science Goes Wrong" the Economist or the title "botch in Science" by Ranga Yogeshwars WDR program " quarks & Co. ".

You may go to the account of this system and the exponential increase in the total retractions. Exponential curves are often run in nature on feedback mechanisms back. If now for about a generation of scientists established researchers to teach the young how to sell unfinished or poorly designed experiments to top journals in order to get hold of a chair, rather than to make solid, well-controlled and reproducible work, one can easily imagine how it can lead to an exponential increase in the retraction. Incidentally, this exponential already reached in 2046 a value of 100% of retracted articles - currently we stand at only 0.02%. If already on 0.02% of retracted articles to reports in the media about the lack of reliability of publicly funded science, one can see that this exponential curve will end well before 2046. It is up to us whether we control the shape of the curve, or whether ultimately the outraged taxpayers rightly decides that the research billions are wasted and otherwise better to use.

A third important aspect of our infrastructure are of course their costs. As for the source code, there is no institutional and only a small part of the information infrastructure that we have to limit ourselves in this discussion on the cost of the literature. Here, the analyst of Outsell, Inc. have calculated that worldwide about U.S. $ 10 billion a year is spent on subscriptions scientific literature. In approximately 2 million published articles per year, this means that any subscription paid items U.S. $ 5,000 costs. This is much higher than even the most expensive Gold OA fees, which are between two and three thousand U.S. $ and more than ten times the average Gold OA fees. Therefore, even a little desirable for other reasons, conversion to Gold OA appears at least financially very lucrative.

Scientific articles at a price of U.S. $ 5,000, the piece can still afford a few institutions, most of which

now for over 15 years, an alternative system that provides the same basic services as the traditional, peer-reviewed journals, but at significantly lower costs and fully open to the public. The Scientific Electronic Library Online, or SciELO, was developed in Brazil with support from the NIH / PubMed, now serves all of South America and begins to expand to other continents. The newest member of SciELO is South Africa. The average cost per item in SciELO amount to U.S. $ 90. In other words, would the scientific world by publishing collective in SciELO starting tomorrow, not only all access problems would be solved, we would also save 9.8 billion U.S. $ annually, without having lost any functionality even in the smallest. And this money could be found for example in turn, could be used directly in the development of digital infrastructure that is worthy of a 21st century.

If all the lives that could save the free access to biomedical research potential, not money saved to be with the nearly 10 billion U.S. dollars every year incentive enough for the development of a modern information infrastructure? If not, then the scientific community has the disaster that just rolls toward her, really deserves.

Björn Brembs is Professor of Neurogenetics at the Institute of Zoology at the University of Regensburg and avowed proponent of Open Access and Open Science.

Last Modified: 07/11/2014