Plan I - Towards a sustainable research information infrastructure

Authors: Björn Brembs[1], Konrad Förstner[2], Michael Goedicke[3], Uwe Konrad[4], Klaus Wannemacher[5], Jürgen Kett[6]
1. Universität Regensburg
2. ZBMED – Informationszentrum Biowissenschaften
3. Universität Düsseldorf
4. Helmholtz-Zentrum Dresden-Rossendorf
5. HIS-Institut für Hochschulentwicklung
6. Deutsche Nationalbibliothek

Introduction

Public institutions in many countries are required by law ("spending rules") to initiate a bidding/tender process above a certain procurement threshold. Scholarly journals are exempt from these spending rules, because the content of each journal can only be obtained from a single publisher - the "single source procurement" exemption. One consequence of this publisher monopoly are prices ranging 10-20 fold above publishing costs [1], or difficult and drawn-out negotiations to achieve technically trivial improvements (such as, e.g., improved accessibility, 'open access'). This "vendor lock-in" prevents marked-based price pressure and stifles innovation. Therefore, functionalities such as efficient citation linking, interactive data visualizations or interoperabilities with data and code have yet to be implemented in the scholarly literature despite sometimes decades of scientist demands. The European Commission (DG Competition) has also acknowledged this problem [2]. For future services, concerning research data and scientific source code there is still a possibility to prevent such a vendor lock-in, with all its detrimental consequences, but time is of the essence. In this document, we would like to elaborate on the problem and provide suggestions for solutions. We see our suggestions as alternatives to "Plan S" or "Plan U", so we chose to name our suggestion "Plan I", for infrastructure.

Problems

While public institutions hesitantly deliberate, commercial providers act

About 30 years after the start of the digital age, even the institutions of higher education around the globe are beginning to consider a modernization of their digital research infrastructures - an embarrassing situation for researchers, tasked by the taxpayer to innovate. In parallel to these considerations the commercial enterprises have already begun to broaden their lucrative monopolies to encompass the entire research information infrastructure. In the course of these efforts, the big international publishing houses have all either acquired or developed a range of services aiming to cover the entire scientific process from literature search, to data acquisition, analysis, writing, publishing and outreach (Fig. 1). In parallel, these corporations have implemented tracking technologies from the commercial internet which they use to survey researchers in high resolution not only to open up new revenue streams by selling that data [3], but also to use the data in-house to be able to offer bespoke packaged workflow solutions (Fig. 1) to institutions.
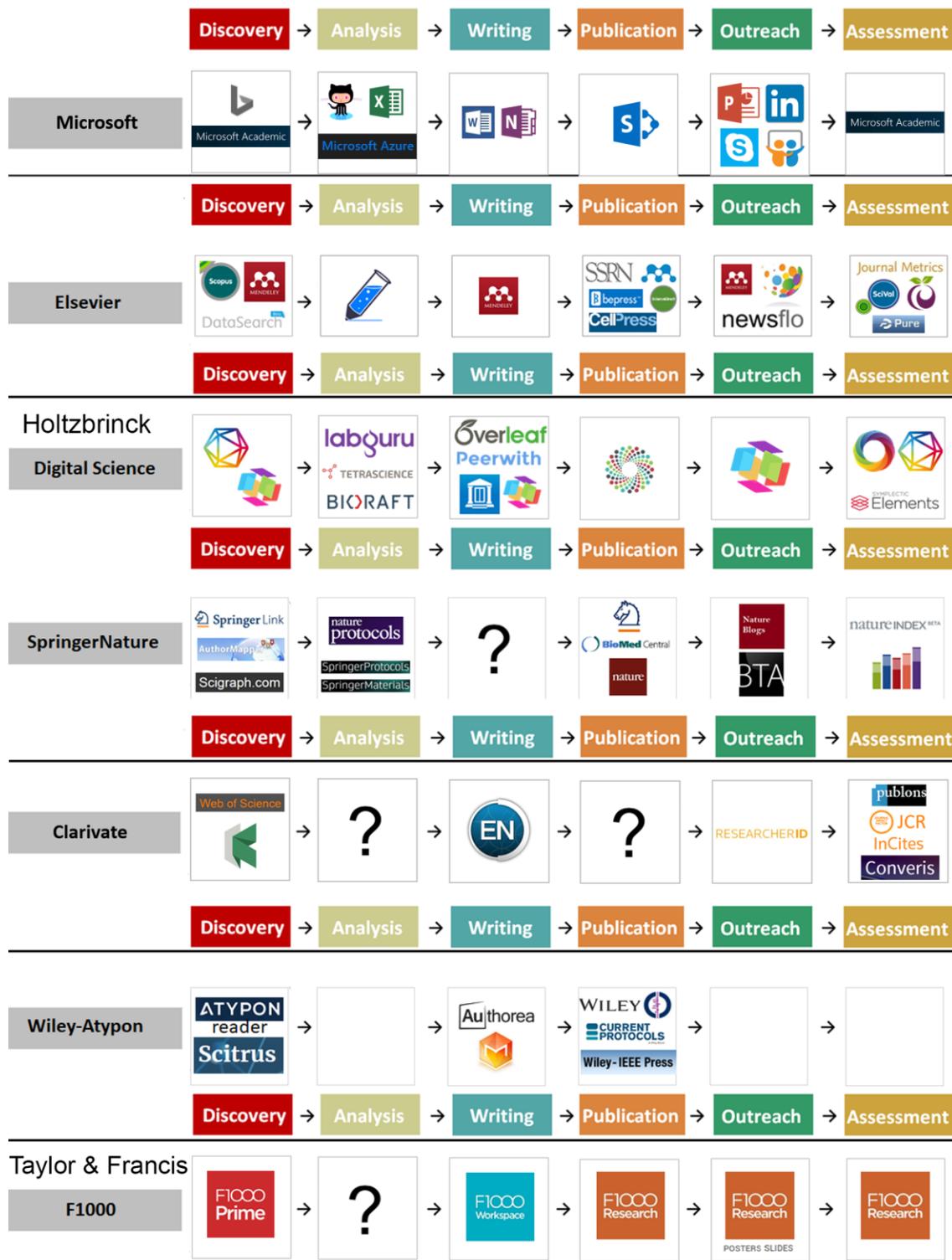
Fig. 1: Providers of digital tools for the scientific workflow (CC BY: Bianca Kramer, Jeroen Bosman, https://101innovations.wordpress.com/workflows). The preconditions for a functioning market exist, but a common standard is missing that provides for the substitutability of service providers or tools.

Vendor Lock-in

Just as in legacy journals, the risk of vendor lock-in also in the area of the new tools and services supporting science is very real: without open, independent standards, it becomes technically and financially nearly impossible to substitute a chosen service provider with another one. In the best case, this non-substitutability will lead to a practically irreversible balkanization of research outputs as long as a plurality of service providers would be maintained. In the worst case, it will lead to complete dependence of a single, dominant commercial provider. A prime example for such a worst-case scenario already in place in academia are electronic laboratory notebooks (ELNs) used by experimentalists in the natural sciences. While there are many commercial providers for such ELNs, open standards that would ensure that the content of each ELN can be effortlessly and cheaply transferred between ELNs is missing. This constellation entails that the data deposited in such ELNs become hostages of the providers who can, consequently, dictate their conditions.

Collective action

Because of the analogous situation with scholarly journals, the danger of a complete monopolization of the scientific workflow becomes more and more likely as time goes by. Their obscene profits have allowed publishers to pick the best tools from the market and incorporate them into their portfolio (Fig. 1), allowing them to now offer institutions nearly complete packages. At the same time, these profits provide them with huge budgets for marketing, lobbying and legal action. This advanced state of development of the commercial providers necessitates collective action on the side of public institutions in the next few years to ensure the substitutability of service providers and prevent a monopolization as in scholarly journals.

Unified goals, balkanized initiatives

The central, overarching goal in implementing a digital information infrastructure for research must of course be to enable all researchers world-wide to work effectively with their data and code, in order to create knowledge in text form or other narrative formats. This goal requires international coordination for integrating data, code and narratives. However, observing the current heterogeneous landscape of approaches, e.g. Plan S or DEAL for narratives, the European Open Science Cloud (EOSC) for data and GitHub or institutional repositories for code, the suspicion becomes inescapable that there is no awareness of these aspects all constituting components of the same infrastructure problem. Moreover, splitting the task between different organizations hampers the necessary, quick collective action (one part of the larger collective action problem) and reduces the effectiveness of the individual initiatives.

Suggestions for a solution

Open standards as a prerequisite for substitutable providers

Open standards, according to which text, data and code will be developed, worked with, made accessible and archived allow public institutions to develop tender or bidding processes, in which service providers can compete with each other with their services for the scientific workflow (Fig. 1). The criteria defined in these standards (i.e., following Open Science and FAIR principles) not only allow for the substitutability of service providers but also assist scientists in following the guidelines for good scientific practice. Such standards thus prevent vendor lock-in, increase price pressure, promote innovation and increase the reliability of science.

Consistent encouragement

Unfortunately, the main reason why now, after 30 years, even before Covid-19, many institutions have started discussing research data management, is not the belated insight that researchers are generating valuable data, but rather the requirement by funding agencies to provide research data management plans and sustainable data archiving and re-use. Similarly reluctantly, many institutions have, in a hurry, invested in new infrastructure components only after social distancing rules prevented in-person interactions on campus. Apparently, it takes binding requirements or global catastrophes to break the inertia of our institutions and bring about essential and long overdue modernizations. Binding guidelines by funding agencies, such as those for good scientific practice by the German DFG [4] are a formidable tool to encourage the receiving institutions to finally spring to action. Developing and implementing open standards would go a long way towards preventing vendor lock-in and support researchers in their work. Funding agencies can incentivize such developments, e.g., by only considering applications from researchers at institutions that have already implemented such standards. Funding agencies insisting that funded institutions cease using counter-productive evaluation criteria such as publications in certain journals by publicly signing initiatives such as DORA, would also support decreasing the market power of legacy publishers.

A source of funding for the implementation

Subscriptions of scholarly journals currently cost public institutions about tenfold more than the actually accruing publication costs, largely due to a) obscene publisher profits, b) inefficiencies and c) financing of non-publication costs [1]. This entails that an alternative publication system in which the current monopolies were replaced by a market characterized by substitutable service providers would stand to save our institutions 90% of the currently spent subscription moneys. The European Commission has already demanded such substitutability and has explicitly pointed out H2020/ERC, Plan S and EOSC to take steps to enforce it [2]. This alternative system would neither disrupt the publication of scholarly articles nor threaten access to the already published literature. With the EU open access publishing platform Open Research Europe (ORE) becoming part of Open Research Central (ORC), there is already a framework available for every institution to move towards such a system with substitutable service

providers. Plan I would entail, e.g., that funding agencies require every institution to join this platform before considering any applications.

Conclusion

Research and scholarship are crucially dependent on an information infrastructure that treats all scholarly output, text, data and code, equally and that is based on open standards and open markets. With concerted action it is possible to realize such an infrastructure without additional costs to the scientific community. The benefit to society, due to the increase in efficiency and reliability of science, would be enormous. Researchers, decision-makers and civic society must work cooperatively and quickly towards such a solution.

References

[1] https://f1000research.com/articles/10-20/v1

[2] https://zenodo.org/record/2565052/

[3] https://zeitung.faz.net/faz/geisteswissenschaften/2020-12-02/3a45541d53e23d3ae2cbea2c3e6a0af1/?GEPC=s9

[4] https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/