BIBLIOTHÈQUE SCIENTIFIQUE NUMÉRIQUE
DIGITAL SCIENTIFIC LIBRARY

Critical study of the new ways of “editorialising” open access scientific journals

Synthesised report of the study piloted by BSN 4 and BSN 7
Produced by Pierre-Carl Langlais (June 2016)
Steering committee: Serge Bauin, Emmanuelle Corne, Jacques Lafait, Pierre Mounier

Introduction

As this report goes to press, France has just adopted an open access law. Article 30 of the "Law for a Digital Republic" provides that the authors of scientific texts whose work is at least 50% publicly funded may "make [it] freely available in an open format, through digital means" after a period of restricted access (known as an "embargo" period) lasting no longer than six months in the technical and medical sciences, and twelve months in the humanities and social sciences.

In September 2016, this measure was definitively approved by the National Assembly and the Senate. It aims to remove one of the main restrictions limiting the dissemination of scientific publications on digital networks: the need to obtain the publisher's authorisation. The exclusivity clauses typically included in publishing contracts potentially prevent authors from re-disseminating their scientific contributions (for example, in an institutional repository) or else apply varying restrictions.[1] These complex arrangements do not facilitate the implementation of a coherent open access policy, neither for authors (who must adopt a radically distinct republication strategy, depending on the publisher) nor for readers. The embedding of open access in the law creates a minimum harmonised framework capable of simplifying the conditions attached to accessing publicly funded research.

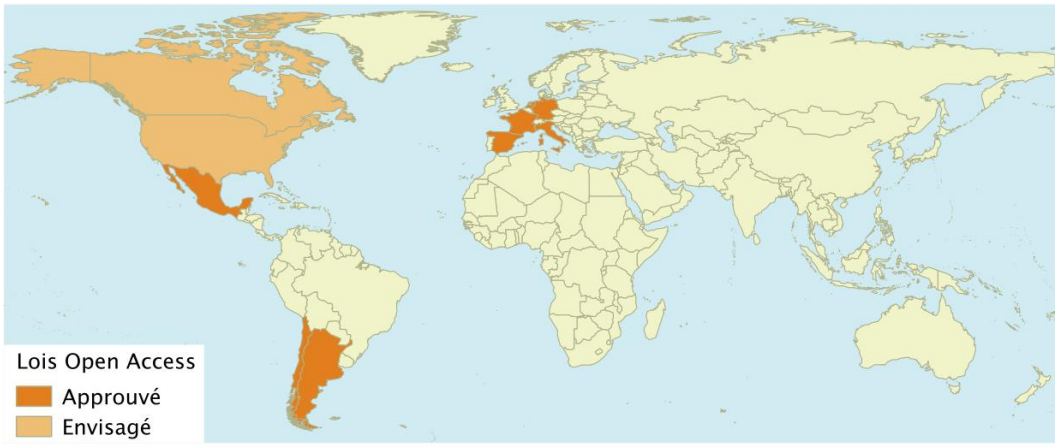
The French law builds on a European, indeed global, movement. On 17 July 2012, the European Commission put forward a recommendation on access to and preservation of scientific information.[2] Based on the assumption that access to publicly funded research contributes to "speed[ing up] scientific progress", this text advocates that such research should be disseminated "as soon as possible, preferably immediately, and in any case no later than 6 months after the date of publication, and 12 months for social sciences and humanities" (Art. 1).[3] Four years later, on 17 May 2016, the Council of the European Union recommending rolling out this model to all Member States in order to make open access a "default option by 2020".

---

[1] See the conditions listed by SHERPA/RoMEO: http://www.sherpa.ac.uk/romeo/journalbrowse.php

[2] Recommendation 2012/417/EU: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012H0417

[3] Council Conclusions on "the transition to an Open Science system", http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf



Map of approved and planned open access laws

In this transition from contract to law, the practice of open access entails a major redefinition of the existing business models. Scientific publishing remains overwhelmingly structured around a subscription-based distribution system, essentially targeted at libraries and scientific institutions. Several French and European reports have attempted to assess the feasibility of a conversion to open access. Odile Contat and Anne-Solweig Gremillet's recent study for the BSN, *Publier : à quel prix ?* [Publishing: At what price?], highlights the extent to which French humanities and social science journals are dependent on public funding: the "vast majority of publishers associated with journal-producing organisations are essentially publicly funded".[4]

Today, the debate is moving from an assessment of the principle of open access towards a more specific discussion of the possible models. The question is no longer "Should we move to open access?", but rather: "What type of open access do we want?"

Certain major initiatives within the open access movement focus exclusively on the managerial issues involved in converting scientific publishers' distribution model. The twelfth conference of the Berlin Declaration organised in December 2015 by the Max Planck Gesellschaft thus led to a proposal to "flip subscription-based journals" by transferring the funds allocated to subscriptions to covering "processing costs":

[4] Odile Contat and Anne-Solweig Gremillet, "Publier : à quel prix ? Étude sur la structuration des coûts de publication pour les revues françaises en SHS", *Revue française des sciences de l'information et de la communicatio*n, 7, 2015, https://rfsic.revues.org/1716

> The objective of the conference was to build a consensus for an internationally coordinated effort to shift libraries' journal budgets away from subscriptions and towards [an] article processing costs model for open access journals (via APCs).[5]

This focus on converting business models overlooks a major sticking point: the review process. Some bodies or university communities do not recognise open access publications as scientific contributions. Section 60 of the French National Council of Universities thus provides that "Articles in solely 'open access' journals will not be taken into account".[6] Paradoxically, although the reliability of open access publications is disputed in the name of standardised review protocols, these same protocols are themselves increasingly challenged.[7] Open access could thus draw its legitimacy from a potential counter-model of the review process, one that would no longer be based on the implementation of closed procedures conducted in private between publishers, authors and reviewers: "open peer review".

Remit of this report

This report was initially commissioned as part of a twofold critical assessment of the new forms of open access publications and those that preceded them: what publishing forms can the state encourage in a digital age that is witnessing the transformation of scientific publishing and the failure of scientific peer review?

The shift from one publication ecosystem to another not only raises a series of "challenges" (digital technology, changing publishing practices, etc.), it also brings to light pre-existing flaws that are otherwise generally hidden by the inertia of habit: the increase in the number of scientific corpuses accessible online allows for a fine-grained appraisal of the deficiencies of peer review protocols, while the overhaul of the conditions in which scientific

---

[5] "Berlin 12 Conference Focuses on Proposal to Flip Subscription Journals to Open Access", *arl.org*, http://www.arl.org/storage/documents/publications/MaxPlanckBrief-March2016-1.pdf, accessed 15 November 2016.

[6] http://www.cpcnu.fr/web/section-60/conseils-generaux

[7] This discourse of crisis crops up frequently in journal editorials. See, for example, A. Mulligan, "Is peer review in crisis?", *Oral Oncology*, 41(2), 2005, p. 135–141, for a discussion of how the revelation of serious fraud has discredited the process ("These events have acted as a catalyst within the scholarly community, with many questioning the role of peer review"), or more recently, Gottfried Schatz, "The faces of Big Science", *Nature Reviews Molecular Cell Biology*, 15(6), 2014, p. 423–426, for a discussion of the negative effect of regrettably low acceptance rates ("Science needs competition, but competition has become so fierce that many fields of science now resemble war zones. Nothing illustrates this better than the current crisis of peer review").

texts are brought into circulation incidentally encourages experimentation in other types of publishing activities.

At this critical juncture where the established norms are becoming more fragile and the norms of the future remain vague, institutions and scientific communities find themselves in a position to influence the reconfigurations underway and encourage long-desired innovations. The principle of the reproducibility or openness of research data is not consubstantial with open access dissemination, but its institutionalisation is facilitated in a context of widespread change. The three parallel stages specified in the initial remit of this report depict an overall configuration in which the development of a public policy to incentivise certain models or practices is becoming a necessity. These are, namely, the acceleration of the process of informatisation (above and beyond formats such as the PDF, which faithfully reproduces the appearance of a print edition); the redefinition of business and publishing models (is there still a need to talk in terms of journals, or even articles?); and the critical rereading of the efficacy of review mechanisms (an issue that relates not only to peer review itself, but also to all ways in which peer review is reviewed, e.g. metrics, list of qualifying journals, etc.).

What concrete forms could such an incentivising policy take? In an ecosystem as "interdependent" as digital scientific publishing, it is difficult to imagine encouraging a few innovations on a selective and exclusive basis. The term "innovation" usually refers to the delimited integration of a new product or practice in an entrepreneurial setting.[8] The development of open forms of review or alternative metrics (or, in the end, the conversion to all-out open access) does not fit into this restrictive definition: it is not an isolatable "product", but the redefinition of a fundamental aspect of an activity.

The spread of open access would require less an "innovation policy" and more an "infrastructural policy", one which, as well as supporting uses and specific tools, would go further and define the convergent linkages between mechanisms, actors and practices. A recent report from Knowledge Exchange thus recommends "putting down roots" by embedding open access in a network of actors and convergent policies: "The fundamental challenge for the implementation of OA policies is the need to develop a fully functioning OA

[8] See, for example, the definition put forward by Benoît Godin in "'Innovation Studies': The Invention of a Specialty", *Minerva*, 50(4), 2012, p. 397–421.

infrastructure from the current disparate collection of services."[9] Web publication lends itself to this approach: texts are immediately inserted into an intertextual network of hyperlinks, recurring formats (defined by style sheets or transclusions) and, above all, standardised information flows. Application programming interfaces (APIs) automate the retrieval of metadata and thereby ensure cross-linkages between activities or distinct dimensions of editorial tasks. In concrete terms, distinct services are able to "converse" with one other by immediately establishing standards that define and delimit certain types of information.

This report is structured around the main "fields" mentioned in the initial question. The first section considers the transformation of publishing systems (Editorial Manager, Open Journal Systems) into autonomous "platforms" connected to a whole set of external services. The second section is devoted to the diversification of the forms of scientific contributions, which manifests itself in the fragmentation of publication across several writing spaces (data and programme repositories, etc.) and the emergence of fundamentally hybrid forms, such as code notebooks. The third section considers the current flaws in the standard peer review system (particularly in terms of reproducibility) and the various alternatives that are taking shape under the collective banner of "open peer review". The fourth section covers the reconfigurations of business models: economies of scale generated by the informatisation of editorial procedures, the diversification of resources, and the development of not-for-profit outfits.

All these changes are closely connected to one another. A fifth, concluding section evokes the necessary shift from a policy of innovation, centred on targeted change, to an infrastructural policy: the priority now is no longer to simply label desirable forms of innovation but rather to identify the frameworks most conducive to the harmonised integration of the diverse aspirations emerging in scientific communities. The evolution of scientific publishing could thus include the construction of structured ecosystems, thereby ensuring that texts and information are relayed from one actor to another.

This type of study requires a composite approach. The different aspects of a textual production represent convenient fields of observation rather than autonomous entities. Indeed, the study of editorialisation must endeavour to describe their points of convergence,

[9] *Putting Down Roots*, JISC, p. 5, http://repository.jisc.ac.uk/6269/10/final-KE-Report-V5.1-20JAN2016.pdf

their irreducible concomitance within writing systems that are being constantly redefined.[10] Applied to scientific publishing, the study of editorialisation encourages us to treat the editorial impact of open access seriously: altering the conditions of circulation will necessarily have an impact on the textual device as a whole.

To pursue this composite approach, we have developed three distinct methodologies:

- Building a database containing an "emblematic" selection of publishing organisations and journals: This selection includes journal publishers (Elsevier, PLOS, BMJ, EDP, PeerJ, Nature MacMillan, Biomed Central), learned societies (APS, ACS, European Geosciences Union), experimental projects (RIO, Self-Journal of Science, The Winnower, *Polymath*, F1000, Episciences, *ReScience*) as well as a few "intermediary" actors (the publishing tools Editorial Manager, Open Journal Systems and Scholar One). We have endeavoured to maintain a diversity of publishing models above and beyond the standard distinction between "traditional" and "experimental" journals and restricted or open access. The database as a whole is first and foremost a tool to aid reflection. By forcing ourselves to systematically fill out certain predetermined categories, we ensured that we observed most of the forms assumed by each organisation or journal's editorialisation system.
- A series of interviews with staff working in these organisations: We gave priority to identifying "invisible" elements in online editorial policies – in the first instance, the longer-term strategies and positions adopted by actors in the face of expected innovations. This collection of interviews enabled us to identify the dynamics and tensions and thus to extrapolate changes and short-term strategic shifts.
- The construction and reuse of complementary datasets: In effect, our selection says little about the extent of the change or the effective adoption of these embryonic standards within the 50,000 scientific journals indexed by CrossRef. We used text mining methods to delineate certain intellectual debates (for example, on open peer review). We also retrieved and analysed several datasets relating to, for example, the use of tools and publication forms in scientific research.

Part 1: Publishing tools

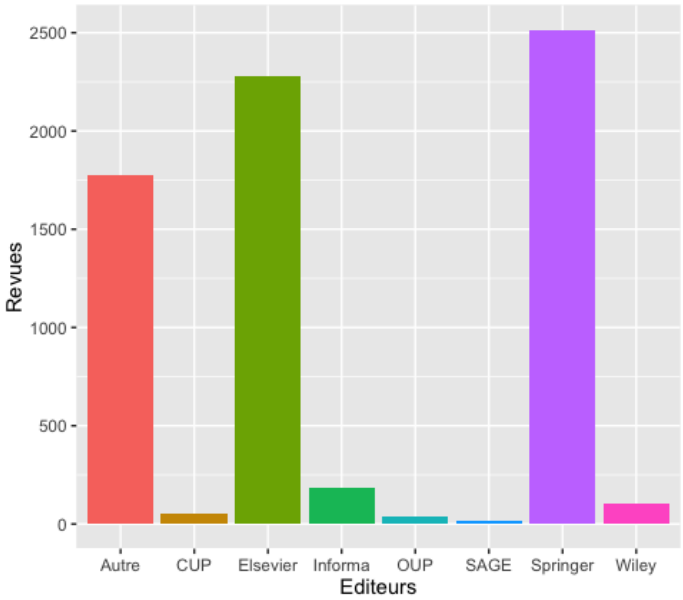
Until the 2000s, the software infrastructures used by journals and repositories of digitised articles were as a general rule based on products that had been developed in-house. Over

[10] Here we borrow the provisional definition put forward by Bruno Bachimont in "Nouvelles tendances applicatives : de l'indexation à l'éditorialisation", in *L'indexation multimédia*, Paris, Hermès, 2007.

the last fifteen years, these "homemade" systems have gradually been replaced by standardised publishing tools, of which Scholar One is perhaps the oldest. Currently owned by Thomson-Reuters, it is used by over 3,000 journals. Editorial Manager has established itself as the dominant actor in this market (with almost 7,000 publications). Open Journal Systems (OJS) is free software produced by a non-profit organisation, Public Knowledge Project (it is difficult to pinpoint the precise number of users, which fluctuates between 9,000 and 24,000). The open archives situation, for its part, is rather different: open source tools such as DSpace predominate (and are used about six times more than proprietary solutions such as Digital Commons).

With the exception of OJS, these tools are mainly used by journals owned by large organisations. In Editorial Manager, only one in four journals is not produced by a major actor in the publishing market. Elsevier and Springer are particularly predominant in this distribution: these two organisations have more journals to their name than all the other clients. The low number of "small" publishers can be explained by the high fees commanded by a premium product. These over-the-counter transactions are not made public, but it is thought that each year PLOS pays almost €7 million to Editorial Manager, i.e. approximately 15% of its turnover.[11] It is easier for large or medium-sized organisations to make this investment profitable, since they can take advantage of economies of scale by converting several hundred or even thousands of journals to a standardised publishing system.

[11] Figure put forward by Andrew Kern after cross-checking several accountancy sources: https://storify.com/jtth/andrew-kern-on-plos



Distribution of digital journals using Editorial Manager (by publisher)[12]

Publishing tools as a "platform"

Commercial publishing tools are more reminiscent of editorialisation systems than "exportable" tools: they do not simply equip pre-existing structures but constitute, in part, a new structure determining the conditions through which information is accessed and texts circulate. Here we borrow the definition of a "platform" put forward in a recent report from the National Digital Council (CNNum):

> A platform is a service fulfilling an intermediary function in access to information, content, services or goods edited and provided by third parties. In addition to its technical interface, it organises and hierarchises content in order to present and connect it to end users. This common characteristic is occasionally accompanied by an ecosystemic dimension characterised by interrelations between convergent services.[13]

The three main criteria identified by the CNNum apply to both Editorial Manager and Scholar One:

- An intermediary service: In concrete terms, "platformisation" translates into the direct hosting of publications and intermediary texts deposited by authors, reviewers and editors. On Editorial Manager, the admin area is not located on the

[12] This distribution by journal does not take into account the number of articles published: a mega-journal like *PLOS One* thus has the same weight as an Elsevier or Springer journal.
[13] CNNUm, *Digital Ambition Report*, June 2015, p. 395, https://contribuez.cnnumerique.fr/sites/default/files/media/CNNum--rapport-ambition-numerique.pdf

journal's site but in a subfolder at the website root. Scholar One, Editorial Manager's main competitor, does not seem to have taken the plunge as far as direct hosting is concerned. The publishing tool does however have a substitute, namely full access to Web of Science, which is published by its parent company, Thomson-Reuters.

- Automation of editorial decision-making: The activity of publishing tools is increasingly based on a form of social engineering that automates in particular the interactions that precede the peer review process. Editorial Manager and Scholar One's current commercial strategy is partly based on the development of features to "automatically suggest" reviewers ("reviewer location"). The vocabulary and the frequency of the terms employed in a submitted article are compared to a vast corpus of articles in order to identify potential reviewers working on similar subjects. The editor makes the final choice, based on the suggested selection. Thus, the editorial decision that is the choice of reviewer is delegated to algorithms, which has an impact on the morphology of the scientific publication: as on Facebook or Twitter, it is based on the selective processing of textual data and social "graphs" (networks of citation or collaboration).
- Ecosystemic dimension: Publishing tools no longer consider the journal as a unit of reference but as an interconnected space. For Josh Dahl, the strategy currently pursued by Scholar One consists in "becoming integrated into an ecosystem", particularly by improving the "connectivity of services" (for example via the development of standardised APIs): "We can longer operate in a vacuum, independently of what other actors are doing and the new open standards."[14] In these circumstances, the publishing tool is no longer simply a writing system (which conditions the morphology of the scientific text), but also a documentary infrastructure that defines the modes of circulation of textual productions. As a stakeholder in an ecosystem, the publishing tool is directly involved in a collective negotiation of standards: Editorial Manager has thus updated its classification of editorial roles in close cooperation with the CREDIT project and the ORCID initiative.[15] If publishing tools believe they can free themselves from journals, the reverse is equally true. The main publishers are now seeking to "repatriate" the editorial process to tools that have been developed in-house. PLOS is currently devoting a large proportion of its surplus resources to the development of Aperta, an "open source" editorial system destined to supersede Editorial Manager.

This conversion of tool into platform derives from the specific constraints linked to the development of scientific contributions. Peer review most resembles a complex system of

[14] Interview with Josh Dahl.
[15] Interview with Mark Hester.

correspondence between actors, which encourages recourse to algorithmic regulatory processes. The ecosystemic dimension derives from the intrinsic "connectivity" of research texts, which are designed to form part of citation and indexing networks, potentially unified by cultural norms.

Open source software: Tools without platforms?

Over the last fifteen years, open source software has established itself as a serious alternative to proprietary solutions. Used by an estimated 10,000 journals, if not more, Open Journal Systems is ahead of its competitors Scholar One and Editorial Manager in absolute terms. In open archives, the predominance of open source tools is even more pronounced: they account for over 60% of the use of such tools (the market leader, DSpace, represents 43%), while the figure for proprietary tools is 10% (OpenDOAR data).

Open source tools seem to be at a remove from the tools-becoming-platforms dynamic. In Open Journal Systems, mechanisms to retrieve data and texts are non-existent, and the software designers do not know exactly how many users there are. For want of being integrated into a uniform infrastructure, the software's ergonomics remain problematic. By continuing to be an "exportable" tool, OJS must be directly "customised" by users to suit their needs. Technical adaptations remain necessary. The technical documentation is sometimes inadequate: the help page describing the ways the tool can be customised is "very patchy and sometimes difficult to follow".[16]

In the absence of direct revenue, open source tools depend on institutional support: OJS was initiated by the Public Knowledge Project, and DSpace by MIT. Yet improvements to the software are mainly dependent on the work of employees: "Most development work is performed by the professional team of developers at the Public Knowledge Project, even if, by the very open source nature of the tool, we can rely on the input of a few outside contributors." Securing this institutional support over the long term can sometimes prove problematic. The DOAJ free software file lists fifteen in total, but most seem to have become

[16] Jean-Luc Archimbaud, synthesis paper on OJS, deposited on *archivesic* at the following address: http://archivesic.ccsd.cnrs.fr/sic_01074813v2/

inactive or to have disappeared altogether: the links to topaz, peerlibrary, hjournal and gapwork sent us to non-existent servers.[17]

Open source tools now tend to diversify their partnerships, in particular with institutions and European projects (such as OpenAIRE). Despite these difficulties, open source publishing software has contributed to reducing the design costs of a large number of journals by disseminating automated procedures that have long been applied within large organisations such as Elsevier or Springer.

Recommendations

The complex development of scientific publishing tools makes it possible to achieve substantial economies of scale; in this respect, these tools contribute to the expansion of publishing and business models for scientific journals. Nevertheless, scientific communities and institutions do not make sufficient use of the opportunities (as well as the risks) that these tools present, which translates into an even smaller volume of contributions to free software such as OJS.

This observation leads us to make the two following recommendations:

1. Encourage scientific communities and institutions to become involved in developing these tools. This point is essential: it will determine the academic world's capacity to develop its own writing practices without being conditioned by outside influences.
2. Take responsibility, either directly or indirectly, for part of the development and maintenance of open source tools. These tools contribute to democratising the economies of scale achieved by automation and the simplification of procedures. The cooperation between OJS and OpenAIRE shows that support can take the form of regular collaborations on projects or with institutions and university libraries, and then progress to "in-kind" subsidies (through the participation of developers associated with these organisations).

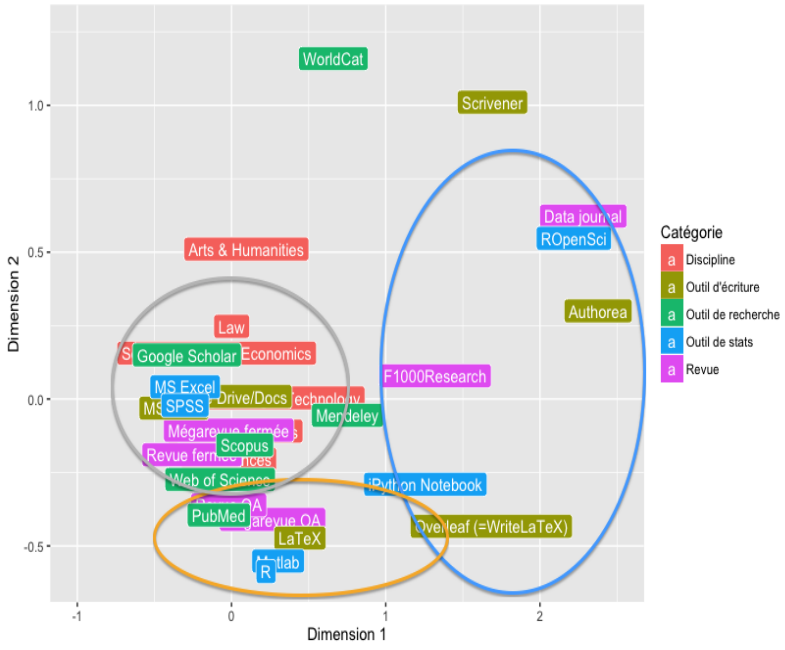
Part 2: Publication forms

In 2001, Tim Berners-Lee predicted that the research article would soon be a thing of the past on the "semantic Web", arguing that it will be possible to publish experimental results directly

[17] http://oad.simmons.edu/oadwiki/Tools_for_OA

on the Web itself, outside the confines of the research article.[18] Fifteen years later, this promised revolution is yet to materialise. The vast majority of scientific texts are either still published in a format that is very reminiscent of print (the PDF) or in basic HTML architecture (which, like most webpages, distinguishes only the most basic editorial structures: paragraph, table, title, etc.).

Despite the inertia of practices, cultural conceptions of the scientific text have not remained inert. We performed a correspondence analysis on the results of a survey of 20,000 researchers across the world, conducted in 2015 and 2016.[19] In the graph below, several coherent sets can be observed: publication in "closed" journals goes hand in hand with the use of proprietary tools (Word, SPSS, Excel) and/or tools owned by the multinationals of scientific publishing (Web of Science, Scopus, Mendeley), while open access publication is more compatible with the use of free software (R and Latex, although MATLAB is a notable exception). A third set, which is equally dominated by free software, seems to correspond to a new demand for "reproducibility".



Correspondence analysis of the use of tools and infrastructures, with the following sets: closed publication (grey), open access publication (orange), and reproducibility (blue)

[18] Tim Berners-Lee and James Hendler, "Publishing on the semantic web", *Nature*, 410(6832), 2001, p. 1023–1024.

[19] For a presentation of the study, see Bianca Kramer and Jeroen Bosman, "Innovations in scholarly communication: Global survey on research tool usage", *F1000Research*, 5, 2016, http://f1000research.com/articles/5-692/v1

This mapping of usages seems to suggest a partial shattering of the paradigm that currently prevails. Although the revolution promised by Berners-Lee has not yet materialised, new "editorialisation systems" based on the coherent interrelation of tools, documentary material and cultural conceptions of the text are nonetheless emerging.

Publishing data

Scientific journals have long published data series: historically, in some disciplines such as astronomy, compilations of tables even precede the publication of research articles.[20] In the context of a digital edition, the verbatim reproduction of tables that were designed for a print publication produces suboptimal results: it imposes limits in terms of the number of entries and columns and does not lend itself to future reuse (in contrast with specific formats such as the CSV).

Data publication today increasingly involves an intermediary: the data repository. Publishers' servers are generally not optimised to manage and disseminate high-volume compilations. Several actors have established themselves in this niche, such as Data Dryad (a not-for-profit organisation), Figshare (a business) and Zenodo (an infrastructure developed by CERN and OpenAIRE). While Zenodo is free, Data Dryad and Figware's business model is based on the receipt of data processing charges, which are a kind of hybrid between hosting costs (an investment that has to be renewed each year) and article processing charges (these actors do not simply store data but also offer editorial features such as the indexing of datasets).

Although the organisations surveyed are delegating the editorialisation of data to intermediaries, it seems they are not entirely abandoning this activity. PLOS and PeerJ have made the cross-publication of data and texts the default rule, and any exceptions must be justified (for example, if sensitive information is involved). Elsevier is reorienting its publishing strategy in this direction, taking advantage of the experience of its parent company, RELX, in managing big data. As with publishing tools, the repatriation of this "subcontracting" cannot be ruled out in the short to mid term.

[20] Catherine Vassilieff, "Les revues françaises d'astronomie. Un cas d'interactivité scientifique", *Revue de synthèse*, 135(2–3), p. 239–254.

Whether they are hosted by a repository or a journal, scientific data retain a peripheral status: supplementary materials that support a pre-existing article. The recent development of "signs of legitimisation" (the assignment of a DOI, or the calculation of metrics) has occurred under the auspices of a reproduction of the documentary device that is the article. Datasets have the status of closed collections: Figshare and Zenodo's search engines index metadata only, and not content per se. Set reading pathways making consulting the article a necessary first step before accessing the data. Data journals are not fully liberated from this standard framework: they are more reminiscent of method journals, emphasising the conditions in which datasets are collected and analysed.

The lack of interrelation and interoperability between data today presents a whole host of issues that extend well beyond the specifics of scientific publishing. The lack of infrastructures to facilitate the identification and retrieval/use of pre-existing data results in an increasing degree of "duplication", with several departments or companies independently performing similar work. The "semanticisation" of data remedies this dissipation of human and technical resources. The standardised labelling of relations and categories makes it possible to directly interrogate existing data without having to identify and download a pre-existing collection.

Today, this dynamic of semanticisation is being initiated by new actors: innovative organisations such as OpenEdition and self-governed communities like the "free knowledge base" Wikidata.

In 2013, OpenEdition rolled out the use of Text Encoding Initiative (TEI) schemas. This programme of standardising online textual formats also includes a large number of attributes that make it possible to "natively" semanticise scientific outputs: standardised tags indicate that a statement relates to a certain date, place or person. Thus the Hellenic era can be defined in a machine-readable form: "<date notBefore="-0323" notAfter="-0031">from the death of Alexander to the Roman conquest</date>".[21] This effort to semanticise is also pursued by institutional initiatives such as the TGIR Huma-Num and its Isidore platform, which are designed to collect and connect research data.

[21] Example taken from http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ND.html

Like Wikipedia, Wikidata emanates from its contributors, who not only edit its "content" but also decide on its morphology more generally. The categories used are subject to open and constant discussion and suggestions. This continued feedback on standards and uses extends Wikidata's coverage to the whole range of knowledge by developing and systematising appropriate fields for every knowledge form: it is thus possible to specify that historical data is hypothetical or was valid only during a certain period in time.

Generating composite forms

Will the new importance attached to distinct forms of the written-up text (data, IT programmes, interactive modelling, etc.) entail a profound transformation of the ways in which science is written? For the time being, the widespread adoption of the PDF format means that there is a neat continuity with the stylistic uses inherited from print journals. The debate over the "reproducibility" of methods and results nevertheless encourages the introduction of experimental mechanisms that closely interconnect text, data, code, video and interactive models.

The journal *ReScience* is hosted on the platform GitHub, which specialises in collecting computer programmes. Its founders, Nicolas Rougier and Konrad Hinsen, aim to "respond to the lack of replication of results" by developing a writing mechanism that ascribes a central importance to the "code archive":

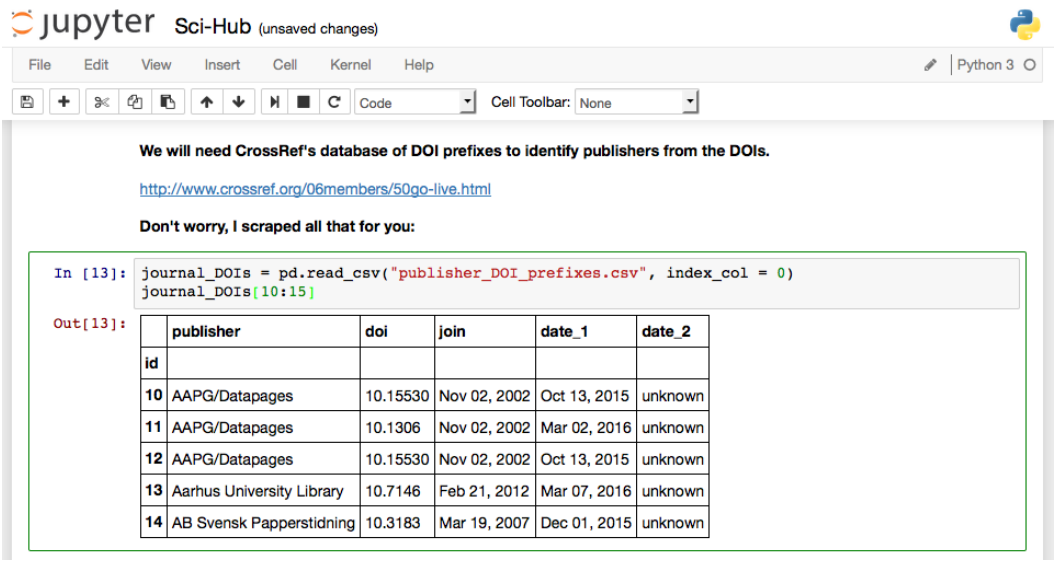
> [T]he editorial chain is radically different from that of other traditional journals. *ReScience* is based on GitHub, where each new implementation is made available with commentaries, explanations and tests.[22]

On the face of it, the standard notion of the article is no longer relevant in this context. *ReScience*'s experimental efforts extend beyond the mere co-publication of code. The written text "splinters" into a multitude of distinct forms: comments, interactions, codes, data. This innovative practice generates "traditional" articles. The written-up text remains preponderant, as do its time-honoured subsections ("Introduction", "Methodology, "Result"). Nicolas Rougier acknowledges that the current situation is not satisfactory. Now that the editorial procedures have stabilised, he would like to see the format of articles evolve: "In the end, the publication standard will no longer be the PDF but a truly digital format that will not simply

[22] Interview with Nicolas Rougier.

attempt to mimic print."[23] He considers an emerging format, the code notebook, as a means to completely extricate ourselves from the usages established in scientific journals over nearly half a century.

The first notebooks of this kind were introduced in 2011 by the IPython project, which in 2015 became the Jupyter project.[24] IPython/Jupyter's approach is the exact opposite of the one pursued by traditional journals attempting to branch into the co-publication of code: the code itself is the reference writing form, and is gradually enriched by incorporating more elaborate texts than straightforward "comments".



	publisher	doi	join	date_1	date_2
id					
10	AAPG/Datapages	10.15530	Nov 02, 2002	Oct 13, 2015	unknown
11	AAPG/Datapages	10.1306	Nov 02, 2002	Mar 02, 2016	unknown
12	AAPG/Datapages	10.15530	Nov 02, 2002	Oct 13, 2015	unknown
13	Aarhus University Library	10.7146	Feb 21, 2012	Mar 07, 2016	unknown
14	AB Svensk Papperstidning	10.3183	Mar 19, 2007	Dec 01, 2015	unknown

Code notebook opened (and modified) in the Jupyter application[25]

In comparison with other usages that are common to all computerised writing forms, the "comment" fulfils a preeminent function: its role is not simply to explain the choices implemented in each function, but also to guide the line of reasoning (via notations in bold type). Once the reader has installed the IPython application, he or she can directly modify the

[23] *Ibid*.
[24] Cyrille Rossant mentions the possibility of taking inspiration from "proprietary" projects like Mathematica or Maple. See Cyrille Rossant, *Learning IPython for Interactive Computing and Data Visualization*, Packt Publishing Ltd, 2015, p. 2. The creation of Jupyter marks an expansion of the project beyond the Python programming language (the name is a portmanteau word, combining the languages **Ju**lia, **Pyt**hon and **R**).
[25] Code notebook used for the following article: John Bohannon, "Who's downloading pirated papers? Everyone", *Science*, 2016, http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone

code and therefore alter the output result. In the image shown above, we altered the original code: it no longer displays the file header, but rather lines 10 to 15.

In the short term, code notebooks will be able to establish themselves as a key format in the computer sciences. And elsewhere? In the study by Kramer and Boesman cited above, the use of code notebooks is fairly significant, given that it is a technology that was created only a few years ago and is still at the experimental stage: 5.18% of those surveyed used this technology regularly. The highest adoption rate was found not in the engineering and technological sciences (including computer science) but in the physical sciences, with adoption rates of 8.36% and 9.72% respectively. Use of this technology remains more limited in the human sciences, but is not entirely anecdotal (1.5% in the HSS, 1.3% in law,[26] 0% in the arts and humanities).

Converting forms: From text to data, and vice versa …

The development of text and data mining techniques marks another convergence point between text, codes and data. The issue here is not to make these composite forms coexist but rather to "convert" them, ensuring that the text can be transformed into a series of structured data and, inversely, that these data can once again be presented to the user in the form of a written-up text.

The automated extraction of texts exploits the lexical or syntactic recurrences of natural language to transpose the knowledge described in standardised and structured informational architectures. The scientific text lends itself particularly well to this exercise: over decades, if not centuries, of collective negotiations, communities of researchers have come to agree on a set of "standards", i.e. terms or expressions designating a certain object or relation as unambiguously as possible.

Text and data mining techniques save a significant amount of time. Thus, in the space of a few months, the Text2Genome project has been able to retrieve the genomic data accumulated in several million scientific articles.[27] These possibilities have recently begun to attract the interest of the major projects focused on the compilation and curation of scientific

[26] In the field of law, the representativeness of the sample is uncertain: only 7 users of code notebooks were identified among 500 researchers (versus 99 users among 6,465 researchers in the HSS).
[27] http://bergmanlab.ls.manchester.ac.uk/text2genome/

data. A text mining application, StrepHit, will soon be extracting the data from scientific publications and relaying them directly onto Wikidata.[28] In this case, the circulation of the text-turned-data does not stop there: Wikidata entries enable the automatic generation of articles in languages that are less common on Wikipedia (for example, most African languages have fewer than a thousand articles). The ubiquity of scientific knowledge is complete: the written-up statement becomes structured data which, in turn, can generate another written-up statement.

This new mobility of data and texts raises several major challenges. The legal framework is ill-adapted. While the practice of text and data mining aims only to extract "raw facts", it implies the dissemination of copies of the analysed databases and corpuses among members of the project, in other words, outside the "family circle". Several exceptions relating to text and data mining for research purposes have been introduced in the last few years in the UK, Japan and France (as part of the Digital Bill), as well as in the United States (through case law). Relaxing the legal framework appears to be a precondition for the achievement of sustainable projects: the introduction of the copyright exception in the UK was immediately followed by the launch of one of the current main initiatives in this area, ContentMine, which aspires to collect several hundred million "facts" from the scientific literature.

Given the current state of the infrastructures, the extensive circulation of text–data and data–text conversions may harm the traceability of information. In contrast with the frequent use of complete references in scientific articles, the schemas of digitised databases generally do not include a specific field to indicate the original source. This feature exists on Wikidata but remains underused. [29] In May 2016, a new initiative, WikiCite, launched with the ambition of developing tools and standards suited to the citation of sources on Wikidata and online databases more generally.[30]

Recommendations

Several convergent changes attest to a structural transformation of the practices of scientific writing. The production of data and computer programmes appears to be less and less a

[28] https://meta.wikimedia.org/wiki/Grants:IEG/StrepHit:_Wikidata_Statements_Validation_via_References
[29] Around 20% of the statements are referenced to external sources: https://tools.wmflabs.org/wikidata-todo/stats.php
[30] Presentation of WikiCite: https://meta.wikimedia.org/wiki/Wikicite

peripheral and hidden activity: dedicated publishing spaces (data repositories, data journals, etc.) and hybrid forms (code notebooks) are contributing to their editorialisation. These different incarnations of a scientific output tend to coincide: text and data mining techniques thus ensure the reciprocal conversion of texts and data.

The standards currently in place do not seem well adjusted to these emerging practices. This leads us to make the following recommendations:

- Introduce a copyright exception for text and data mining: Reforms seeking to secure the practice of TDM are currently being discussed in several European countries. Only a universal copyright exception for research purposes will effectively remove these restrictions.
- Initiate and support the process of semanticising data: Too often, research data remain isolated. Given the current state of the infrastructures, it is difficult to identify and retrieve/use completed research. Semanticisation makes it possible to label data in accordance with interoperable standards, which could be developed collaboratively by scientific communities.

Part 3: Peer review

The open access movement was born of a crisis: the *serial crisis*, which was characterised by the exponential increase in journal fees. It is now contributing to revealing another: the *replication crisis*.

This crisis is a direct consequence of immediate and unrestricted access to scientific articles on the Internet. The partial dissolution of disciplinary frontiers (hitherto shored up by preferential subscription policies practised by libraries), and, more generally, the possible intervention of an "initiated" audience situated outside of academia, are both contributing to multiplying the number of potential critical perspectives. New bodies are conducting *post hoc* reviews that are resolutely distinct from those performed by journals: contributors to the anonymous forum PubPeer have detected several cases of previously unnoticed fraud (including the case of the French biologist Olivier Voinnet last year).

These verifications demonstrate that many articles are not reproducible: by strictly replicating the same method and similar data, attempts at reproduction culminate in different results. As a consequence, the retraction rate (i.e. the *post hoc* rejection of an article that had

previously been approved, generally after a flaw in the peer review) has noticeably increased.[31]

Open forms of review have been enjoying unprecedented popularity over the last few years. The Google Scholar statistics illustrate the increasing weight of this notion in the scientific debate: since 2010, the number of results has been following an exponential curve, with more than 1,000 publications indexed in 2015. Beyond experimental trials, leading actors are beginning to branch into this field: in the short term, PLOS is seriously considering rolling out open peer review. Elsevier is also carrying out pilot experiments in this area. Agnès Henri of EDP Sciences points out, however, that the development of appropriate mechanisms raises numerous questions: "Editorial committees are quite enthusiastic in principle; as soon as discussion turns to implementing concrete procedures, they tend to become much more wary. Should we publishing everything? And in what form? … The issue of identity disclosure also raises problems: should the documents be anonymous or not?"[32] There is no universal model for open peer review, but rather a family of models based on the adjustment of several factors. We will consider the three principal factors in turn: the visibility of the author; the morphology of the writing device; and the later circulation of reviews.

The contributors

Should the author be made public, just like his or her contribution? This is both an epistemological and sociological issue: anonymity seems to have an impact on the reviewer's involvement. PeerJ was eventually won over to a system of voluntary open peer review, which includes the possibility of an anonymous contribution. According to Jason Hoyt, "we didn't want to scare away authors or potential reviewers". In the same way, the PubPeer team reveals that anonymity has played a decisive role in the expansion of the community: "Comments on PubPeer have been hugely encouraged after a user-controlled anonymity system was put in place."[33]

[31] R. Grant Steen, Arturo Casadevall and Ferric C. Fang, "Why has the number of scientific retractions increased?", *PloS One*, 8(7), 2013, p. e68397.
[32] Interview with Agnès Henri.
[33] "Vigilante Scientist", *blog.pubpeer.com*, http://blog.pubpeer.com/?p=200. Quite fittingly, the post is also anonymous.

Some forms of anonymous open review have already been trialled on a considerable scale outside the field of scientific publication. The 100,000 regular contributors to the various Wikipedia communities favour the use of anonymous accounts or pseudonyms in order to focus the encyclopaedia's debates on the ideas put forward, rather than people. Although it facilitates the involvement of reviewers, the practice of anonymity could also be more easily misappropriated.[34]

A second demarcation line relates to the way in which contributors are recruited. In practice, all the forms of open peer review that we have been able to study limit involvement in the review process to authors of scientific contributions. In the experiment conducted by the journal *VertigO*, this limitation was not deliberate: the journal has "invited its readers to comment on the texts, but [the approach] has not delivered results".[35] *Polymath* also sought to be open in principle, but has mainly gathered comments from experts. The development of sustainable linkages between participatory science and open peer review remains hypothetical.

The barrier to entry varies. On PubPeer or the Self-Journal of Science, the entire scientific community is able to review any article, without the upstream involvement of an editor: the researcher simply has to want to participate.[36] Other journals maintain a mechanism for selecting reviews that may revolve around an editor (this is notably the case at *PeerJ* or the *British Medical Journal*) or authors themselves (authors are strongly encouraged to suggest reviewers on F1000). The *VertigO* experiment concluded that it is useful to maintain, at least temporarily, a form of editorial intermediation: targeted letters addressed to one reviewer in particular have met with much greater success than general announcements.[37]

[34] There have been several cases of hidden reviews on PubPeer (for example, one author passed himself off as a peer, giving himself a good review): https://forbetterscience.wordpress.com/2015/12/13/post-publication-peer-review-signed-or-anonymous/

[35] Julien Bordier, "Évaluation ouverte par les pairs : de l'expérimentation à la modélisation", 2016, p. 33. Uploaded to HAL at the following address: https://hal.archives-ouvertes.fr/hal-01283582/document

[36] The definition of "scientific community" varies between the two initiatives: on PubPeer, it includes all authors indexed in the PubMed and ArXiv open archives; on Self-Journal of Science, membership is dependent on affiliation to a scientific institution.

[37] Julien Bordier, *op. cit.*, p. 33.

Forms of peer review

Peer review is now implemented in a great variety of ways, aiming not only to disclose the texts that are exchanged but also, potentially, to alter them completely. This brings us back to the terms of our initial question on the editorial issues associated with open access: in editorialisation systems, all "dimensions" of textual production are closely correlated.

The first attempts at open peer review have not totally extricated themselves from the stylistic practices followed by journals: one of the pioneers of this practice, the *British Medical Journal*, confines itself to an act of transparency by publishing the email exchanges. New actors are attempting to introduce new approaches.

The Self-Journal of Science publishes the review process in real time: the articles submitted are immediately published, then reviewers leave their remarks as comments; an article is officially accepted once it has received the approval of five members. This approach requires no editor to act as the intermediary between the authors and reviewers, since the latter put themselves forward of their own accord.

The Hypothes.is project has developed an annotation system for the Web that is now being integrated into several scientific publishing tools (including Open Journal Systems). Peer review does not simply take the form of a unified "text", but is rather a continuous series of comments regarding the scientific contribution. The review does not simply derive from the reviewer's expertise, but from an argument-based and continued exchange.[38]

The Self-Journal of Science and the VertigO experiment represent complete editorial solutions. Several open peer review actors are positioned solely in the review sector itself. We have already mentioned the PubPeer forum: RetractionWatch also fulfils much the same function (*post hoc* verification of articles that have already been approved). Publons and Rubriq, for their part, are more focused on the division of labour: they deliver a rapid and efficient review of the submitted proposals (in Rubriq's case, by remunerating reviewers), and then send them to partner journals.

[38] *Ibid.*, p. 30.

Encouraging the circulation of open peer review

Since the 1960s, the circulation of reviews has been confined to the calculation of metrics. At the institutional level, the assessment of the quality of a scientific output or journal is frequently based on the calculation of the citation rate or "impact factor". Over the last dozen or so years, this practice has been called into question by several alternatives: altmetrics. The most common altmetrics cover the number of downloads and unique visitors, as well as the number of times an article is mentioned on social networks. Nevertheless, the development of alternative indicators remains hindered by the lack of a universal open access citation index: although Google Scholar can be freely consulted, there is no API to automatically retrieve this information.[39] Several ongoing projects aim to collate the citations of open access publications (since there is nothing to stop them being freely retrieved/used). Since early 2016 Science Open has been working on an Open Citation Index, which currently includes 2 million articles "mainly extracted from PubMed and ArXiv".[40]

An alternative exists: once opened up to the public, peer review becomes a mobile text, circulating in the same way as metrics. Like scientific articles, it can be indexed or disseminated by other publication forms (bibliographic indexes such as CrossRef, open archives, academic social networks). PeerJ has recently begun assigning a DOI to reviews and indicating their citation rate. Taking into account new issues such as reproducibility contributes to valorising peer reviews and almost confers on them the status of "secondary" scientific contributions. In 2015, the PubPeer site introduced an add-in for the browser Firefox, one that made it possible to find comments associated with each scientific contribution indexed in its database. On Self-Journal of Science, curation itself constitutes a review practice: each researcher can create his or her own log by collating the articles judged emblematic in a research field. A thoughtful selection serves as an endorsement and provides a *post hoc* validation of a scientific approach: "In contrast with citation, reuse is much more difficult to manipulate: it is a much stronger editorial act and one that is immediately visible."

[39] See our attempt to retrieve data from Google Scholar using scraping (which quickly led to the programme being blocked): "Faire du data mining avec Google : comme tromper big brother ?", http://scoms.hypotheses.org/216

[40] Jon Tennant, "The Open Citation Index", *scienceopen.com*, 29 February 2016, http://blog.scienceopen.com/2016/02/the-open-citation-index/

The circulation of reviews attracts the interest of open archives. The OpenAIRE project has recently developed an Open Peer Review Module (OPRM), an add-in connected to an open archive and making it possible to directly deposit reviews without having to go through a journal or intermediary (such as PubPeer).

All these connections allow for the concomitant existence of several review systems. The evaluation of any one article on Self-Journal of Science, the OPRM of an open archive, or an open and anonymous forum will probably lead to different and complementary conclusions. For example, to the extent that it is open to the general public, PubPeer seems to lend itself more to the detection of errors and statistical anomalies than to the overall assessment of the pertinence of a study within its specialised research field. The proliferation of review forms opens up the prospect of an enriched meta-peer review, one able to respond to the new expectations associated with the reproducibility of research.

Recommendations

As with publication forms, review practices are increasingly out of step with the current standards. The proliferation of critical points of view on scientific contributions, beyond peer review, is already underway: interactive forums such as PubPeer are revealing pre-existing flaws, while the different models of open peer review implemented over the last few years appear to be viable alternatives, to the extent that established publishers (PLOS and EDP Sciences) envisage deploying them.

- Rethink the review of peer review: The administrative provisions that regulate research usually reduce the set of possible review forms to the standardised peer view protocol implemented by leading journals. For example, the AERES has produced a list of qualifying journals for the humanities and social sciences. This qualification does not take into account new intermediary actors (such as Publons or Episciences) that cover only part of the journal's editorial work.
- Highlight the various forms of review in open archives: New editorial mechanisms could specify that such and such an article has been certified by a journal or by an independent review body. This development would contribute as much to valorising open archive platforms (which would become the central convergence point for different forms of review) as it would to facilitating the advent of new actors (whose capacity to certify a scientific contribution would be recognised from the outset).

Part 4: The business models

One of the largest scientific libraries in the world today is actually a "pirate" website. Sci-Hub hosts and disseminates over 50 million articles, for the most part obtained illegally. The size of the corpus is comparable to the (legal) collections of metadata that are Web of Science (90 million indexed articles) or Scopus (55 million articles). The development of Sci-Hub is based on a kind of industrialisation of pirating, marked by the absorption of pre-existing biblioleaks[41] and the automation of "leaks": when an article requested is not already available in its database, Sci-Hub will automatically attempt to obtain it from one of the publisher's platforms.

The recent arrival of this new actor has contributed to "dramatising" the question of the future of scientific publishing's business models. Based in Kazakhstan, Sci-Hub has little chance of being closed down by a court decision and can count on the tacit or explicit support of "authors".[42] The convergence of the needs of research "user-authors" and the demands of institutions implies a general short-term shift. We will examine the main aspects of this transformation in turn: the evolution of publishing costs, the diversification of resources, and the development of not-for-profit ventures.

Publishing costs

Currently, the cost of a journal is closely linked to its publishing model. On the Web, the cost of dissemination is gradually becoming too low to invoice (or "too cheap to metre"). With CERN's colossal storage capacity at its disposal, the data repository Zenodo is completely free of charge (unlike its main "competitors", Figshare and Data Dryad). According to Laurent Romary, much the same also applies to HAL: "[The cost of] hosting an article amounts to €5 and this cost item is continuing to fall as the platform is optimised."[43] In short, the pooling of costs proves to be collectively less expensive that the ascription of a price (and the associated bureaucratic infrastructure that would be required to manage these transfers of funds).

[41] See Guillaume Cabanac's in-depth research on this subject: "Bibliogifts in LibGen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing", *Journal of the Association for Information Science and Technology*, 67(4), 2016, p. 874–884.
[42] See the survey on the website of the journal *Science*: http://www.sciencemag.org/news/2016/05/survey-most-give-thumbs-pirated-papers
[43] Interview with Laurent Romary.

Publishing tools also benefit from this economy of scale: whether Open Journal Systems is downloaded by hundreds or tens of thousands of journals, this produces almost no additional cost for its designers.

This strategy of reducing costs nevertheless runs into a major difficulty: the need to adapt equipment and tools to the specific needs of the journal and a certain research community. Computer journals have a significant advantage here, in that they can rely on the direct participation of those who initiated them. The team at the *Journal of Machine Learning Software* or *ReScience* ensure that the publishing equipment is constantly maintained and adjusted; here, technique is never dissociated from practice. According to Kevin Stranack (of Open Journal Systems), recruiting developers within libraries and scientific institutions could contribute to "democratising" this important interrelationship. These developers could in particular design add-ins that are directly tailored to the needs of users.

In the interviews, peer review was frequently seen as an "incompressible cost". Though the review itself is conducted on a voluntary basis, putting authors and reviewers into contact can represent several hours, or even days, of work. An editor at the American Physical Society observes that a "truly selective and editorialised journal cannot be profit-making: it loses money on the articles it refuses".[44]

Would it be possible to reduce the costs associated with peer review, for example, by partially automating the procedures? This prospect is currently inspiring various experiments (such as the automated recognition of reviewers, the optimisation of interactions on PeerJ, or the future PLOS publishing system, Aperta). The savings generated by these innovations are significant: PeerJ is able to charge much lower fees than its competitors, for an apparently comparable result (€100 per author). Some experimental open review approaches presented in Part 3 go even further, by completely liberating themselves from the costs of peer review. On *ReScience*, the Self-Journal of Science, the Open Peer Review Module and PubPeer, authors and reviewers interact without the need for an intermediary.

Resources

Only the diversification of resources makes it possible to reconcile open access publication with the maintenance of a minimally profit-making distribution model. Three main

[44] Interview with an APS editor.

possibilities are currently taking shape. They are not closed models but rather seem destined to hybridise.

- The author-pays model: This is by far the most favoured solution at present. The most recent negotiations between Elsevier and national consortia of libraries and scientific institutions thus point to the gradual migration of funds earmarked for subscriptions towards the purchase of rights to publish in open access in journals (article processing charges, APCs). According to the terms of the Dutch agreement, 30% of articles should be published in open access by 2018.[45] The increasing take-up of the author-pays model will probably embed a structural trend in the economy of scientific publishing: the shift from a market of "goods" or "services" to a purely "symbolic" market. Estimating the price of APCs over the long term is very uncertain: in fifteen years, the fees charged by PLOS have not fallen but have actually increased (after the last fee increase, the fees at *PLOS ONE* were $1,495[46]). Conventional publishers charge much higher prices (around $5,000 at *Nature Communications*), and indeed these could rise further: according to Mike Taylor, an article distributed on a subscription basis earns Elsevier $7,000 on average.[47] In a purely symbolic economy, there is nothing to prevent the standard rates from rising considerably: "If people are prepared to pay over $230,000 for a degree from Yale … why wouldn't they be prepared to spend $50,000 on an article that is capable of earning them tenure?"[48]
- The sale of associated data: Large scientific publishers are increasingly turning their attention to the business models developed by social networks. Facebook and Twitter's revenue is based not on subscription fees but on the sale of personal data. Academic social networks such as ResearchGate, Academia and MyScienceWork have succeeded in extending this model to the specific case of scientific publishing. Elsevier is also by and large reorienting itself in this direction. Its recent purchase of Mendeley, a bibliography management service similar to Zotero, and the very recent incorporation of the open archive SSRN are the embryonic beginnings of a new department entirely devoted to the sale of browsing data (which is extended by exclusive partnerships with academic social networks like MyScienceWork[49]). At the same time, the publisher has also invested substantially in the development of tools

[45] See the summary of the agreement: http://vsnu.nl/files/documenten/Domeinen/Onderzoek/Open%20access/QA_OpenAccess_Akkoord_Elsevier_ENG.pdf
[46] http://blogs.plos.org/plos/2015/09/plos-publication-costs-update/
[47] https://svpow.com/2012/07/09/what-does-it-cost-to-publish-a-paper-with-elsevier/
[48] Björn Brembs, "How gold open access may make things worse", 7 April 2016, http://bjoern.brembs.net/2016/04/how-gold-open-access-may-make-things-worse/
[49] https://www.elsevier.com/about/press-releases/science-and-technology/sciencedirect-content-now-available-to-mysciencework-users

to review and monitor research, SciVal[50] and Pure.[51] These have been very widely adopted by scientific institutions. A new and profitable market is thus taking shape before our eyes: the user information retrieved from open archives, bibliography management tools and Elsevier-affiliated social networks come together to improve the efficacy (and thus the profitability) of meta-review instruments.

- Supplementary services: This business model serves to clarify the author-pays model. The act of purchasing no longer relates to a right to publish (and thus to the presumed reputation of the journal), but instead applies to specific services. Those services may be essentially symbolic: for example, The Winnower charges for the purchase of a DOI. OpenEdition has developed specific services for libraries and their users. For example, although content is freely accessible in HTML form, PDFs and ePubs are sold (DRM-free) to libraries and individuals.[52] The development of extraction techniques could also contribute to adding value to these services. Fine-grained TEI indexing – and ensuring that data and information are from the outset part of the ecosystems of exchange on the semantic Web[53] – could also increase the value and attractiveness of these services in the long run.

The not-for-profit approach

Of the three business models compatible with open access publication, only the sale of supplementary services produces managed expenses: the costs correspond to measurable services. For the time being, this is not the most popular option. The Max Planck Institute's proposal of "journal flipping" moves in the direction of a more general take-up of APCs, while, at the same time, major actors such as Elsevier are attempt to repeat the success of the "walled gardens" of the Web (such as Facebook) in the world of science. These possible developments lend credence to an apparently radical rupture: total emancipation from the for-profit model. In this context, scientific publishing would no longer sell anything at all (neither articles, publishing rights, nor personal data); its existence would be based exclusively on volunteering and/or institutional subsidies.

This not-for-profit model is already widely practised. In 2008, a JISC report estimated that for a restricted access digital journal, editorial tasks represented 28% of the real costs of

[50] https://www.elsevier.com/solutions/scival
[51] https://www.elsevier.com/solutions/pure
[52] See, for example, the books offer: http://www.openedition.org/13052
[53] For example, via automated identification tools for publication IDs (DOIs), such as Bilbo: http://maisondesrevues.org/680

producing a scientific article (including the author's and the reviewers' contributions).[54] Yet these editorial tasks are not always performed by an editor: the study by Odile Contat and Anne-Solweig Gremillet shows that French journals are usually subsidised "in kind" by delegating the management of peer review and the correction of articles to academic staff, at a median cost of €1,600 per article.[55] Scientific publishing becomes a "market" only at the very end of the process: the writing of articles – as well as the review and, often, the formatting of those articles – is defrayed by public funds.

In principle, the economies of scale generated by digital publishing systems enable an immediate transition to open access simply by reallocating these hidden subsidies. The median sum of €1,600 identified by Contat and Gremillet is comparable to the costs charged by leading publishers (€1,500 at PLOS), and significantly higher than those incurred by new forms such as *épirevues* ("epi-journals") (according to Laurent Romary, the incompressible work of the Episciences secretariat amounts to €200, with the remainder being either insignificant or absorbed in the long term[56]).

The development of open access has been accompanied by a proliferation of non-profit journals: around two thirds of the titles indexed in the Directory of Open Access Journals charge neither the author nor the reader. Although significant in quantitative terms, these initiatives are only just beginning to enjoy a higher profile: the Public Knowledge Project (which has also developed the freeware Open Journal Systems) recently launched a survey on cooperative models in scientific publishing.[57] A longitudinal study published in *PeerJ* in May 2016 indicates that largely volunteer-based organisations are able to continue in the long term but struggle to reach their full potential: the average number of submitted articles remains comparatively low.[58]

---

[54] Calculated from the data on p. 184, http://www.webarchive.org.uk/wayback/archive/20140614211536/http://www.jisc.ac.uk/media/documents/publications/rpteconomicoapublishing.pdf

[55] Odile Contat and Anne-Solweig Gremillet, "Publier : à quel prix ? Étude sur la structuration des coûts de publication pour les revues françaises en SHS", *Revue française des sciences de l'information et de la communication*, 2015, 7, https://rfsic.revues.org/1716

[56] Interview with Laurent Romary.

[57] http://oa-cooperative.org/

[58] Bo-Christer Björk, Cenyu Shen and Mikael Laakso, "A longitudinal study of independent scholar-published open access journals", *PeerJ*, 4, 2016, p. e1990, https://peerj.com/articles/1990/

Not-for-profit journals face two challenges: as well as not having their own funds, they often receive fewer subsidies. Of the three structures included in our sample, only the *épirevues* at Episciences enjoy direct or indirect support from scientific institutions. Paradoxically, purely voluntary organisations do not have the means to directly seek financial assistance. This situation affects the visibility and sustainability of the publishing venture: for Michaël Bon, the Self-Journal of Science community remains limited due to lack of institutional "endorsement", while Nicholas Rougier of *ReScience* confided to us his concerns for the long-term viability of a journal hosted by GitHub.

Several recent initiatives are striving to make this public support "official". The funding of the Open Library of Humanities (OLH) is based on the receipt of library subscriptions: libraries do not buy publications (which are available in open access) or publishing rights, but instead commit investments that, over time, will make it possible to reduce overall publication fees in the humanities. According to the initiator of OLH, Martin Eve, spreading this expense between several institutions makes the individual cost much lower than a subscription (a little over £2 for 190 institutions).[59]

More generic mechanisms appear necessary to ensure the development of smaller entities. In 2014, the Lemoine report on the digital transformation of the French economy advocated introducing an "open 1%", i.e. a levy applied to profit-making activities at the outset "to encourage businesses to commit to open access projects".[60] In the same way, the funding of the new models of scientific publishing could be ensured by a systematic levy on budgets allocated to subscriptions and APCs. Organisations dependent on volunteers would also be strengthened by the introduction of an individual right to contribute (*droit individuel à la contribution*): each year, a small amount of working time could be dedicated to developing not-for-profit projects.

Recommendations

At the economic level, the increasing take-up of open access is often interpreted as an expansion of the author-pays model. In fact, the diversification of publishing models observed in the previous sections of this report translates into a concomitant diversification

[59] Interview with Martin Eve.
[60] Philippe Lemoine, *La Transformation numérique de l'économie française*, p. 21.

of potential business models. Aside from APCs, emerging actors are developing new organisational forms, such as subscriptions, supplementary services and volunteering. Scientific publishers themselves are anticipating this widening of horizons. Over the long term, data exploitation will perhaps yield more profit for Elsevier than the receipt of APCs.

In this fast-changing context, we put forward the following recommendations:

- Transforming certain expenses into investment: The public budget dedicated to scientific publishing essentially aims to perpetuate the existing system. Journal flipping is akin to a mechanism that "shifts" current expenses (which will, at best, remain constant over the long term). In contrast, the emergence and consolidation of new business and publishing models offers the prospect of an overall and massive reduction of the expenses incurred. This investment could take the form of an innovative 1%: all transactions (subscription renewals, APCs, etc.) would include a small levy to finance new models.
- Developing subsidies that are appropriate for not-for-profit entities: These organisations generally do not have the necessary resources to seek the direct or indirect support their commercial competitors enjoy. The introduction of micro-funding mechanisms with minimal administration and/or the recognition of the individual's "right to contribute" would represent particularly appropriate responses. This approach could link in with an overhaul of how peer review is itself reviewed, as discussed at the end of Part 3, by establishing positive incentives (bonuses, etc.) to publish in non-profit open access journals and take part in those journals' editorial development.

Part 5: From innovation to infrastructure

Two factors: Mobility and diversity

The drastic reduction of the cost of replication and republication on the Web and the removal of legal restrictions in the framework of open access are substantially expanding the potential ways in which content can circulate.

A text is likely to have several different lives. A document deposited in an open archive may evolve into a journal article, potentially leading to other texts being formulated (such as open peer reviews). These, in turn, may generate other types of text. Indeed, the very content of the text is liberated: automated extraction techniques permit the migration of information into structured databases; from there, the information can once again be converted into a written-up text.

Each of these successive improvements could be "connected" retroactively, with archives compiling dispersed peer reviews, attempts at replication or the list of "raw facts" extracted by ContentMine. The illustration below shows a projection of the HAL archive interface, enriched with side panels displaying reviews, replications or various data extractions.

The projected integration of derivative texts in an open archive

Here we can see a fundamental linkage developing between mobility and diversity. This projected integration recognises from the outset the possibility of incorporating several forms of review that do not necessarily have the same impact or utility. In Part 3, we saw that the emerging models of open peer review offer distinct and complementary advantages: although the mechanism at PubPeer has proved its effectiveness in detecting fraud, it is probably unable to assume all the functions of standard peer review. Improving the reliability of review processes will entail the coexistence of several systems and, as such, a multiplication of the possible angles and points of view (which, moreover, it will be possible to read simultaneously thanks to new publishing devices).

The substantive work undertaken by publishing tools over the last fifteen years facilitates this link-up between a plurality of models. The various types of possible add-ins and customisations increasingly correspond to interoperable standards. Each publishing organisation may thus identify the appropriate "cocktail" corresponding to its needs and

ambitions while also remaining permanently connected to the flows by which texts circulate and are disseminated.

Innovation, infrastructure, ecosystems: A realisation

The current structure of scientific publishing is characterised by a kind of entropy of effort. Organisations are obliged to commit what are sometimes substantial resources to perform certain activities for which they lack the necessary in-house expertise or experience. Duplication is rife: the same articles are reviewed several times (generally arriving at the same conclusions), features are reinvented by separate teams, and series of similar data are constructed on several occasions.

The reduction of the overall cost will require a concomitant change: integrating innovation into coherent and structured infrastructures and, secondly, ensuring that the infrastructures in these ecosystems are interconnected. The organisations included in our sample voiced this realisation to us with varying degrees of emphasis. The cofounders of *ReScience* are aware of an "uncomfortable" situation: perpetuating a standard form of the article (where code and data files are merely an appendage), despite the fact that the formal written-up text now plays a diminished role in the review protocol. It therefore seems that there are (approximately) optimal ways of arranging the forms assumed by the different aspects of scientific editorialisation. On the face of it, these arrangements are difficult to perceive: they structure through practice. One expression came up time and time again in the interviews: "work in progress". The Winnower is "currently still in the experimental phase";[61] at PLOS, "the roll-out of open peer review is work in progress". No new organisation claims from the outset that it has found the ideal model to embody a certain kind of desired change.

The transition from innovation to infrastructure depends on more than the isolated realisations of individuals, however numerous they may be. In recent years, the European Union has undertaken a coherent infrastructural policy. The publication policies implemented as part of the Horizon 2020 programme require the deployment of appropriate republication

[61] Interview with Joshua Nicholson.

systems: at the same time, the proliferation of database projects[62] straightaway raises the issue of their interoperability. The concomitant creation of the open meta-archive OpenAIRE and the data deposit Zenodo provides the basis of a European ecosystem of scientific publishing.[63] Publishing tools are closely associated with these initiatives: Open Journal Systems is working "closely with the OpenAIRE project";[64] it was also by collaborating with OpenAIRE that DSpace developed a review module for open archives. More generally, many actors are in the process of developing sustained interrelations with this embryonic kernel: in November 2015, OpenEdition updated its publishing infrastructure to ensure it is entirely harvestable by OpenAIRE,[65] while Zenodo has formed a partnership with GitHub to facilitate the mutual export of data and computer files.[66]

The large scientific publishers remain at a remove from this European movement; they are attempting to develop their own ecosystems. Elsevier's acquisition and development policies fully subscribe to this dynamic. Publishing systems (Pure, and later on, EVISE), review systems (SciVal), search engines (Scopus), bibliography management tools (Mendeley), open archives (SSRN) and no doubt over time academic social networks will all form the initial foundations of a captive ecosystem. As with Facebook, Twitter or Google, the use of these different services will be at least partially free. Elsevier, however, has secured complete mastery of the information flows, and reserves the right to exact a price for these different impact measurements from different scientific institutions.

Governance: Inspiration from the commons?

Elsevier's very recent acquisition of the open archive SSRN – which occurred just as the write-up of this report was entering its final stages – revealed an intrinsic weakness in the open access movement: although the use of open licences prevents the appropriation of texts, it does nothing whatsoever to limit the appropriation of editorialisation and dissemination structures. The issue facing us today is no longer simply that of access, but also the means of

[62] To date, over 800 research infrastructures have been funded by European Commission programmes. See the map on http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=mapri

[63] Other initiatives will soon supplement this ecosystem, such as the European Science Cloud, which is destined to "manage data from different sources": http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[64] Interview with Kevin Stranack.

[65] https://leo.hypotheses.org/12612

[66] https://guides.github.com/activities/citable-code/

accessing and controlling communication channels. Academic social networks have laid the foundations of a large-scale "enclosure" of open access scientific outputs. Users must connect to ResearchGate, Academia and MyScienceWork to consult articles and, in doing so, tacitly agree to enter into a commercial contract, namely ceding personal data in exchange for access to knowledge. And yet the value and precision of data is also increasing with the proliferation of such sources of appropriation. The likes of SSRN and ResearchGate can learn much less by remaining independent than by integrating into a conglomerate the size of Elsevier.

As a consequence, governance becomes decisive. In a rapidly changing context in which innovations have yet to stabilise, the sustainability of a publishing organisation can be measured first and foremost by its ability to adopt appropriate and responsive decision-making processes and effectively involve all stakeholders. Thus, the funding system behind the Open Library of Humanities derives from its internal organisation: the subscriptions paid by 190 libraries and higher education institutions have enabled a collegial management structure to materialise; thanks to the direct payment of subsidies, academic communities retain full control over the editorial process.

Governance forms are more poorly documented than tools or writing practices. The DSpace publishing tool manual is almost 1,000 pages long, and yet there is no manual of possible forms of governance for scientific publishing. Each institution must improvise and often individually recreate practices and methods that exist elsewhere. This lack of documentation also hinders usages. Many contributors to the SSRN archives apparently believed it was a non-profit organisation, and as such one that in principle could not be bought out by Elsevier.

In this respect, new scientific publishing organisations could take inspiration from the experience gained from the knowledge commons. The development of digital technologies has permitted the development of very large and self-governing communities. Wikipedia, OpenStreetMap and Debian are founded on the contributions of several hundred if not several thousand users, who are constantly taking decisions about the framework and the

standards of their participation.[67] These commons are not entirely new entities.[68] Scientific publishing as a whole resembles a hybrid commons (or club), such is the extent to which this field appears dominated by externalities that elude the standard analytical frameworks of the market or state.[69]

The shift to the Web has entailed a quantitative and qualitative leap. Although founded on a system of self-governance, the encyclopaedia Wikipedia has now achieved a scale comparable to a cultural industry (or, in the area of scientific publishing, Elsevier): across all languages combined, almost a hundred thousand contributors keep several dozen million articles permanently up to date. The automation of tasks has played a decisive role in the advent of these industrialised commons: the Wikipedia community can delegate the application of standards to bots (thereby ensuring that inappropriate changes are identified) and thus free up time to take part in collaborative forms of governance. The technical infrastructure emanates directly from the needs and uses of the community, to the extent that it could be described as collaborative work by design.

This precedent shows that other models are possible. The role played by Elsevier could be assumed by large self-governed platforms involving all the stakeholders of scientific publishing: researchers, librarians, tool designers, or even, in a spirit of open science, outside participants. More than technical support, what is lacking today is an international collective dynamic: efforts to innovate are often dispersed among international initiatives and/or lack the means to develop fully.

Recommendations

Although the short-term development of open access appears to be a given, the development of appropriate infrastructures is becoming crucial. This presupposes not only technical and editorial efforts (in terms of the "connection" of scientific outputs), but also a social and "political" undertaking: governance forms are decisive in the management of long-

[67] Although it is not really a manual, the P2P Foundation indexes numerous effective practices found in online communities: http://p2pfoundation.net/Category:Peergovernance

[68] In the work of the Nobel economics prize winner Elinor Ostrom, the term "commons" refers to traditional societies that are not organised around the market or state.

[69] This proximity was suggested in a thought-provoking study by Neyron et al., "Principles for Open Scholarly Infrastructures", *figshare*, February 2015, https://figshare.com/articles/Principles_for_Open_Scholarly_Infrastructures_v1/1314859

term change and the construction of an open and sustainable ecosystem. The three following recommendations reflect this outlook:

- Ensure that the general circulation of texts within the scientific publishing ecosystem is visible: This outcome is already partially achieved. HAL's architecture, for example, is optimised to disseminate information and metadata. Over time, the goal should be to develop universal entry points for the different incarnations of a scientific contribution and the various derivative forms that refer to it.
- Document and certify forms of governance: Studies on this subject have been initiated in Canada and Europe, but much remains to be done. This action can take concrete form in the development of documentation (manuals, feedback, etc.) and standardised certification mechanisms (following the example of a project like Move Commons[70]).
- Encourage an international federative dynamic: Currently, there are only four universal meta-portals that provide access to scientific outputs: Web of Science, Scopus, Google Scholar and the pirate library Sci-Hub. The fundamentally global nature of scientific research requires a global response, namely closer link-up between the main actors in open publishing in the broad sense (publishers, archives, institutions, libraries, communities, etc.) and the development of shared infrastructures under the aegis of collegial and collaborative governance.

Conclusion

The data gathered and presented above provide a response to the initial remit of this report: "What publishing forms can the state encourage in a digital age that is witnessing the transformation of scientific publishing and the failure of scientific peer review?" Unless there is a serious about-turn, open access should soon be widespread: the introduction of laws facilitating the depositing of articles in open archives and, most recently, proactive statements from the European Council point in this direction.

The first four sections of this report sought to take stock of how the main aspects of scientific publishing are currently evolving. This general dynamic of change, triggered by the twofold conversion of academic publications to digital form and open access, is opening up new opportunities. For each of these aspects, we have put forward recommended actions that should be pursued to facilitate the advent and emergence of innovative structures which are adapted to the needs of scientific communities.

[70] http://movecommons.org/fr/preview/

- Publishing tools: The development of standardised systems contributes to optimising the publication process and diversifying publishing models. Freeware, such as Open Journal Systems, is conducive to the deployment of regular feedback between practices and devices (through the ad hoc design of add-ins). To ensure the full development of these not-for-profit tools, we recommend initiating an active policy of direct or indirect public support.
- Writing forms: Scientific writing on the Web resembles less and less a formal written-up text. Data, code and hybrid forms (such as code notebooks) occupy an increasingly important position, which translates into the development of purpose-built publication and distribution channels and forms of legitimisation (such as the ascription of a DOI). The current standards remain at a remove from this recent transformation. We advocate that the administrations responsible for public research support publication forms other than the article, and that the state pursues the development of an appropriate legal framework.
- Peer review: Long confined to isolated experiments, open peer review models have fast been gaining ground in recent years. These alternative forms no longer represent a uniform counter-model but rather a diversified whole, which, depending on how the main parameters are adjusted (author, mechanism, circulation, etc.), produces results that are more or less suited to a given situation. As a result of this diversification, we encourage scientific institutions to rethink how peer review is itself reviewed and to introduce suitable publishing mechanisms to highlight the different forms of review.
- Economics: The "raw" transfer of subscriptions to APCs is neither a desirable nor ineluctable outcome. We call on the public administrations to seize hold of the opportunities opened up by the general shift in business models and turn expenses into investments. A small levy on the expenses incurred (the "innovative 1%") and the pooling of resources will over time enable structural and sustainable savings.

Throughout this report, we have made several recurring observations that transcend whatever particular "angle" is being discussed. Part 5 broadly describes the shift from a policy of innovation to an infrastructural policy. The existing models, even as they diversify, are part of an ecosystemic dynamic: the relations between actors are just as important as what those actors do. The mobility of textual productions within the framework of widened open access (incorporating the opening up of review processes, data, code, etc.) facilitates the emergence of these interrelations. To the extent that it expands the field of technical possibilities, informatisation contributes to reinforcing the impact of the organisational or even political models of scientific publishing: governance becomes decisive. The emergence of digital

commons shows that it is possible to design very large-scale projects, comparable in scope to the main scientific publishers, while being regulated by forms of collaborative and participative governance. The deployment of an international dynamic involving most actors from the opening publishing world has now become a necessity: the development of this burgeoning ecosystem – and the prevention of its appropriation – can probably occur only within an overarching global framework.