HAL open science

Full Disclosure: Open Business Data and the Publisher's Cookbook

Sebastian Nordhoff, Felix Kopecky

▶ To cite this version:

Sebastian Nordhoff, Felix Kopecky. Full Disclosure: Open Business Data and the Publisher's Cookbook. ELPUB 2018, Jun 2018, Toronto, France. 10.4000/proceedings.elpub.2018.8 . hal-01816822

HAL Id: hal-01816822

https://hal.science/hal-01816822v1

Submitted on 15 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Full Disclosure: Open Business Data and the Publisher's Cookbook

Sebastian Nordhoff and Felix Kopecky

Introduction

1 For a long time, the discourse on Open Access has been dominated by the plights of the natural sciences. Skyrocketing journal subscription costs have laid a have burden on library budgets, and skyrocketing profits of the major publishers have raised ethical questions about the funnelling of public funds into shareholders' coffers. The solutions proposed have normally tried to resolve these issues, which are experienced in the natural sciences. For instance, the report by Schimmer *et al.* 2015 explain in detail how the "system" can easily be changed to an author-pays system, since there is enough money in the "system", but at a closer look, the empirical evidence presented centers completely on article processing charges, leaving aside book-length works or non-APC-based models.

Gold, Green and Platinum Open Access

2 There have been criticisms of the Golden Road from the beginning, often advocating the Green Road. This again is modelled on practices in the sciences, where we find a culture of short works of article-length, a general propensity and openness for technical solutions and an acquaintance with parametrized work flows. This is not found in the humanities. Works are longer, techical solutions are less welcome and often met with suspicion, and parametrized work flows are all but absent. The discourse on Open Access up to now has by and large neglected the Other Side of Academia, i.e. the humanities. Generally meager funding means that the Golden Road is not a viable model; and lacking technical affinity among the practicioner population means that the Green Road is not a socially viable model either.

3 As an alternative to the Golden Road and the Green Road, so-called "Platinum Open Access" (or "Diamond Open Access") has been proposed (Haspelmath 2013). Diamond Open Access is dubbed "Publisher pays" by Haspelmath. This is misleading since any publisher has limited funds, which will eventually deplete and so some sources of revenue have to be found. "Community pays" seems to be a better term here. The precise kinds of revenue which can be tapped into by a Platinum Open Access publisher will be detailed further down.

Sustainability and Resilience

4 An important concept for running a publishing platform is *sustainability*. Sustainability means that the project can survive after the initial funding and can meet its running costs by its own revenues without requiring external funds. A more demanding concept is *resilience.* We define resilience as meaning a sustainable enterprise which is additionally guarded against risks of losing its community-based soul to the profit oriented sphere (compare Ottina 2013). An example of a sustainable and resilient project would be Arxiv.org, whereas an example of a sustainable project which proved non-resilient is Social Science Research Network (SSRN), which was sold to Elsevier in 2016.

The Need for Empirical Data

5 Community-based projects face the general challenge that the costs of scientific publications are completely opaque. Publishers give wildly varying quotes as to the costs of papers and books. Scielo publication costs are ~90 USD a paper, while Nature Communications charges 6,000 USD APCs. For books, we find Ubiquity Press with £3,450 and Brill with a quoted 18,500 EUR. Maron *et al.* 2016 find ranges between 15,140 and 129,909 USD as the cost price for books at university presses. All this means that new projects are completely in the dark about the costs they will be facing while setting up shop. A detail of the costs of an actual publisher as a yardstick to measure a new project against is direly needed. Adema & Stone (2017: 79) suggest a "Best practice toolkit", which would help Academic-Led Presses (ALP) in the setting up the publishing platform (also see Ferwerda, Pinter & Stern 2017: 10).

6 Furthermore, the provision of empirically tested-business models is a desideratum, as evidenced by Johnson *et al.* (2017: 65).

7 This is what is being provided by the project "Full disclosure" by Language Science Press.

Language Science Press

8 Language Science Press is a community-based platinum OA publishing initiative for monographs and edited volumes in linguistics. Since its inception in 2014, 60 books have been published, about 120 more are in the pipeline.

9 After initial funding provided by the Deutsche Forschungsgemeinschaft, Language Science Press is now relying on a supporter network of 100 institutions worldwide, which contribute to meeting its running costs.

Open publisher

10 Language Science Press has a strong commitment to openness and is always happy to share. Next to all books being CC-BY, the LaTeX source of all books is available on GitHub; all data and figures are being provide in as raw a format as possible; the LaTeX class is available on CTAN; all software is 100% Open Source. This means that the technical setup could easily be replicated.

11 This being said, a software stack alone is not sufficient to run a publishing house. Next to the technical infrastructure, know-how in the domains of finance and community management is also required. In the context of the project "Full disclosure", Language Science Press releases all relevant business data. This includes downloads per book, sales per book, donations, fees, and institutional memberships as well as the costs for personnel, goods, services, and travel.

Lean publisher

12 Language science press is a lean publisher born digital. A number of legacy cost factors were cut out. These include: warehousing, acquisition editors, paywalling, authentication, billing, royalties, IPR management, marketing, book stands, involved author contracts. The author provides a pdf according to LangSci specifications and releases it under a CC-BY licence. The pdf is made available on the LangSci site and various aggregators, as well as via print-on-demand. PR is done via social media. Acquisition is done via the community network already in place.

Community-based publisher

13 The third principle of Language Science Press is that it is a community-based publisher. There are 20 autonomous series with autonomous editorial boards which see to the quality of publications in their respective subdiscipline (e.g. phonetics or translation studies). Reviewing is done by peers as usual, but also proofreading is done by volunteer researchers (currently 300 enrolled), who are keen on seeing early versions of a manuscript and making the book a better book. This is a kind of "donation of time" instead of a "donation of money", which nevertheless means that the cost of producing the book go down considerably.

A note on terminology

14 When talking about the pros and cons of nuclear power plants, one has to talk about Megawatts, Millisieverts, and risk percentages. There is an established vocabulary for talking about matters relating to nuclear power plants. One can agree or disagree about how many Megawatts the yield will be, how many Millisievert represent a critical dosis, or how high the risk percentage is, but the concept in and of themselves are useful to establish a discourse. That discourse can of course be critiqued, but the using the terms Megawatt, Millisievert or risk percentage does not make you an ardent proponent of nuclear power. To the contrary : in order to militate against nuclear power, you better know those terms. In the same vein, terms such as "revenue stream", "contribution

margin", "depreciation" or more basic "product", "customer" or "value proposition" have clear definitions when talking about business models. A clear and established terminology makes the argumentation more transparent, and it helps shed light on what the critical components of a particular setup are, what aspects are in line with the chosen ethical standards, and what aspects are problematic. Using these terms when describing a setup does not logically entail a support or opposition to capitalism or any particular from thereof. In order to make clear that these terms are used in a technical sense without necessarily endorsing them on a society-wide level, we will put them in scare quotes.

Full disclosure

15 In the project "Full disclosure", Language Science Press shares its business model complemented by granular business data as Open Data. This will allow projects in other disciplines to evaluate in how far the model (or parts thereof) are suitable for their discipline. The business data provided can serve as a yard stick to gauge the attractivity of different "revenue streams" when planning for a post-grant future. These documents centered on financial and legal aspects are complemented by a "cookbook" with step-by-step recipes.

16 The provision of a schematic HowTo squared with actual business data will allow for the empirical evaluation of the proposed recipes and some benchmarking.

17 The availabilty of financial calculations will allow to offset one of the main advantages the traditional publishers have over new projects: experience in handling accountancy matters.

Business model

18 The business model comes in analyses of the business case, the market, the organisation, and funding. For all four aspects, a short general introduction is given. This is followed by a description of the solution adopted by LangSci and an explanation why this particular solution was chosen. This chosen solution is then evaluated in a third part, which highlights where the inital assumptions were correct and where they were completely off the mark, in the hope that other projects will find these lessons learned helpful. Each section closes with a discussion of other solutions which could have been chosen as well, and what changes in the outcome would be expected if they were chosen.

19 A "business model" starts with analyses of the "product", the "customers", and the "value propositions" made to the various groups. One important aspect is that the "customer" in academic publishing is actually no straightforward to pin down. A case can be made that authors, readers, universities, libraries, and the general public are indeed all potential "customers" of Open Access publication platforms in the sense that they derive a "benefit" from the existence of such platforms. This implies that these "customers" can be made to contribute a share of the "benefit" towards the enterprise.

20 The books being published by Language Science Press come from authors of countries all over the world, and are read internationally as well. This means that the "business model" must reflect this international character of the operations. A funding model restricted exclusively to, say, Germany, will be unfair, and it will be hard to convince the

residents of only one country to fund operations which profit the whole world. This international setup has to be taken into account not only for financing, but also for distribution and marketing.

21 When we analyse the "benefits" OA publishing provides, we find that it is not only the readers who profit. Also the authors derive a "benefit" (more prestige, faster publication, greater reach). Furthermore, libraries also do "benefit" via a larger array of books they can provide to their patrons. The state at large also benefits (more cost-effective publishing; no dependency on monopolists), as does society at large (better and faster access to scientific findings).

22 Just as it would be unfair to have one country shoulder the operating costs alone, burdening only one group with the costs (e.g. via "author pays" ' models) is questionable as well. So the question is: how can the "operational costs" be shared among the "stakeholders" in such a way that everybody contributes, and that no one is charged beyond their capacities? And how can that be done without undermining the basic tenets of Open Access?

23 Language Science Press has opted for a broad basis for financing its operations. The diversification of "revenue streams" means not only that each of the various "stakeholders" participates, but it also reduces the risk of "funding gaps" as compared to one large funder. If one "revenue stream" breaks down or does not contribute the expected amount, this can then potentially be compensated by the other streams.

24 The analysis of "clients" above shows that the "product" provided by Language Science Press is complex and can access a variety of "markets" with different "target groups."

25 The basic assumption is that each group of "stakeholders" will have have both egoistic motivations (purchase a book for reading it yourself) as well as altruistic motivations (create societal benefit, ensure free access to the knowledge of humanity, ...). We acknowledge both kinds of motivation as valid.

26 The "business model" is based on five pillars, which are supported by the different audiences:

1. print margins (readers)
2. individual memberships (authors, readers)
3. optional author fees (authors)
4. donations (authors, readers)
5. institutional memberships (institutions)

27 Each channel comes with an "overhead" in administration. For instance, institutional memberships require effort to liaise with libraries, and to get the fiscal setup right. This means that less than 100% of the sum on the invoice will be available for book production. But ideally each channel will provide some "net contribution" towards financing the core task of producing books. An evaluation of the LangSci estimates can be found below in the section "Evaluation".

28 The descriptive part of the business model is complemented by a freely available spread sheet with a 5-year projection, where publishing projects can enter their estimates about projected number of books; time needed for various tasks; flat costs like rent; and estimated sales/author fees/downloads/institutional memberships. With all estimates filled in, the spreadsheet will show whether the project will be sustainable 5 years down the road.

Open Business Data

29 The model is complemented by all business data from Language Science Press in 2017 in *.csv format at https://github.com/langsci/opendata. This shared data resource (possibly called "Accountancy Commons") includes all download figures, all sales figures, all expenses (design, travel, salaries, as far as privacy laws permit). This data can be squared with the original business model and allows its empirical evaluation. It will allow other interested disciplines to estimate the costs in their field, adjusting for particular usages of theirs. For instance, a field with mainly prose text might reduce the typesetting estimates compared to linguistics, while a field with heavy use of photographs will have to add additional budget for this.

Cookbook

30 The final element is the "cookbook" of how to set up a community-based publisher with recipes, timelines, and pitfalls to avoid. This is based on the setup of Language Science Press since 2012. Aspects discussed are the generation and the signalling of:

- prestige;
- governance;
- the creation of a community spirit;
- gamification;
- community proofreading;
- marketing; and
- generating revenue.

Evaluation

31 When the business model was devised in 2015, the revenue listed in Table 1 was assumed for each of the streams. The actual revenue for 2017 is given in that table as well.

Table 1: Revenue as projected and realised in 2017

Revenue stream	Projected	Realised
Print margin	24,000 €	5977.31€
Author fees	48,000€	2,500€
Institutional memberships	56,000€	(85,000€)*
Donations	9,600€	2,500€
Individual memberhships	13,200 €	120€
Context: books	48	26

*Institutional revenue is for 2018

32 Table 1 makes it abundantly clear that the initial estimates were completely off the mark. People buy a lot less printed books than expected, they donate less than expected, and they pay less optional author fees than expected. Even if we adjust for the fact that only half the expected books were produced, the numbers are still below the expectations in all domains, with the exception of institutional memberships. We share our misconceptions here so that other projects can learn from our errors.

33 In 2016, a cooperation with Knowledge Unlatched was started. The idea was to have Knowledge Unlatched do the acquisition of institutional members since they have the relevant expertise and contacts. It became soon clear that this was a more viable revenue stream than any of the others, with considerably less overhead.

34 While the total sum of 120,000€ stated in the business model could not be raised, the number of books was also lower, and the position of an accountant was not filled, leading to 1.0 FTE instead of 1.5 FTE and hence lower personnel costs.

Community Efforts

35 An important aspect which outweighs the financial bean counting is the community participation. Language Science Press has been successful in creating a community of linguists interested in Open Access Publication. This community cares for Language Science Press and wants it to survive and to thrive. There are currently over 900 registered public supporters, 300 community proofreaders and over 100 different authors. Even if the scale of operations has to be adjusted following the revenue situation, there is no doubt that Language Science Press as a project will be resilient enough to continue.

Replication

36 It is hoped that the Cookbook, which will be available for the conference, will inspire other disciplines to set up community-based publishing projects as well, and that the business model and the empirical data will help them see where Language Science Press succeeded, where it failed and how these failures can be avoided in future projects. Ideally, these other projects would also make their assumptions available and offer a critical evaluation of what happened to their plans when they hit reality.

Critical perspectives

37 I would like to close by some critical perspectives on open data. Opening up this data means that we are going one step further towards commodification of science. If a book in linguistics can cost as little as 3,000 €, how could one argue that for a particular publishing project in, say, archeology, you have to budget 25,000€? Obviously, there are huge differences between the two disciplines, but will this be clear to funding bodies? If the person-hours dedicated to a book are disclosed, will this not lead to books being selected as to their requiring less-than-average person-hours in order to have a nice balance sheet? Who would continue doing works like "Tone in Yongning Na" about a small language in China with very complex typesetting when a prose text of 300 pages

will also give you a book to add to your list of deliverables? Weighing these issues, we feel that disclosing the experiences Language Science Press has made at this point in time is more important for the general enterprise of free access to knowledge, but we want to caution against overemphasising the value of metrics in scientific publication.

BIBLIOGRAPHY

References

Adema, Janneke & Graham Stone. (2017). *Changing publishing ecologies: A landscape study of new university presses and academic-led publishing*. http://repository.jisc.ac.uk/6666/1/Changing-publishing-ecologies-report.pdf.

Ferwerda, Eelco, Frances Pinter & Niels Stern. (2017). "A landscape study on open access and monographs: Policies, funding and publishing in eight European countries." *Knowledge Exchange*. https://doi.org/https://doi.org/10.5281/zenodo.815932

Haspelmath, Martin. (2013). *Science publication should be seen as a public service (just like science itself)*. https://www.frank-m-richter.de/freescienceblog/2013/02/14/science-publication-should-be-seen-as-a-public-service-just-like-science-itself/

Johnson, Rob *et al.* (2017). "Towards a competitive and sustainable OA market in Europe—A study of the open access market and policy environment." *Research Consulting.* https://blogs.openaire.eu/wp-content/uploads/2017/03/OA-market-report-28Final-13-March-201729-1.pdf.

Maron, Nancy L. *et al.* (2016). *The costs of publishing monographs: toward a transparent methodology*. https://doi.org/10.18665/sr.276785.

Ottina, David. (2013). "From Sustainable Publishing To Resilient Communications." *TripleC* 11(2): 604–613. https://triple-c.at/index.php/tripleC/article/view/528

Schimmer, Ralf, Kai Karin Geschuhn & Andreas Vogler. (2015). *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access*. http://dx.doi.org/10.17617/1.3

ABSTRACTS

This short paper presents the three main outcomes of the OpenAire project "Full disclosure: replicable strategies for book publications supplemented with empirical data": a fully specified business model; accountacy data; and a "cookbook" containing recipes how to set up a resilient community-based book publisher. The provision of these items available for free reuse will allow other publishing projects to understand, adapt, and modify the community-based model of Language Science Press.

INDEX

Keywords: Open Access, Monographs, business model, Open Data

AUTHORS

SEBASTIAN NORDHOFF

Language Science Press, Germany
sebastian.nordhoff@glottotopia.de
(corresponding author)

FELIX KOPECKY

Language Science Press, Germany
felix.kopecky@langsci-press.org