



Knowledge Exchange



Insights into the Economy of Open Scholarship:

A Collection of Interviews



© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A Collection of Interviews

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interviews dated between: May 2018 and February 2019

DOI: [10.5281/zenodo.2840171](https://doi.org/10.5281/zenodo.2840171)

This report is available under a [CC BY 4.0 licence](https://creativecommons.org/licenses/by/4.0/) creativecommons.org/licenses/by/4.0



Acknowledgements

This report is the effort of many. Thanks to the Knowledge Exchange Task and Finish group with experts from all six Knowledge Exchange countries Anne Sofie Fink Kjeldgaard, Herbert Gruttemeier, Jessica Parland von Essen, Lambert Heller, Laurents Sesink, Leo Mack, Melanie Imming, Minna Ahokas, and Zaza Nadja Herbert-Hansen, to Juliane Kant and Jean-François Nominé who coordinated the Task and Finish group and to the KE Office Bas Cordewener and Sarah James. All have fruitfully contributed to this report, although any mistakes and misreportings are the responsibility of the author.

Contents

| | | | |
|--|-----------|--|-----------|
| Preface | 5 | Interview with Jouni Tuomisto | 38 |
| | | References and relevant links | 43 |
| Introduction | 6 | About Jouni Tuomisto | 44 |
| Shared challenges | 7 | A look into ASAPbio | 46 |
| 1. HR: researchers 'in business' | 7 | About ASAPbio | 47 |
| 2. Commercial versus non-profit: walking the line | 8 | ASAPbio: Business model | 48 |
| 3. Infrastructure: make or buy? | 8 | An interview with Jessica Polka | 49 |
| 4. Licensing: pragmatism | 9 | Creative Commons and PLOS | 51 |
| 5. Sustainability and scalability | 9 | References and relevant links | 53 |
| 6. Marketing: build it and they will come? | 10 | About Jessica Polka | 54 |
| 7. Influence on workflows and the research landscape | 10 | | |
| A look into the Open Library of Humanities | 11 | A look into ScienceOpen | 56 |
| About Open Library of Humanities | 12 | About ScienceOpen | 57 |
| OLH: Business model | 13 | ScienceOpen: Business model | 58 |
| Interview with Martin Paul Eve | 14 | Interview with Stephanie Dawson | 59 |
| References and relevant links | 20 | References and relevant links | 65 |
| About Martin Paul Eve | 21 | About Stephanie Dawson | 66 |
| A look into OpenEdition | 23 | A look into HRČAK | 68 |
| About OpenEdition | 24 | About HR AK | 69 |
| OpenEdition: Business model | 25 | HR AK: Business model | 70 |
| Interview with Pierre Mounier | 26 | An interview with Jadranka Stojanovski | 71 |
| References and relevant links | 32 | References and relevant links | 76 |
| About Pierre Mounier | 33 | About Jadranka Stojanovski | 77 |
| A look into Opasnet | 35 | A look into Helsinki University Press | 79 |
| About Opasnet | 36 | About Helsinki University Press (HUP) | 80 |
| Opasnet: Business model | 37 | HUP: Business model | 81 |
| | | Interview with Leena Kaakinen | 82 |

Contents continued

| | | | |
|---|------------|--------------------------------|------------|
| References and relevant links | 86 | Appendix 1: methodology | 126 |
| About Leena Kaakinen | 87 | Participants | 126 |
| | | Interview questions | 126 |
| | | Procedure | 127 |
| A look into Impactstory | 89 | | |
| About Impactstory | 90 | Appendix 2: glossary | 128 |
| Impactstory: Business model | 91 | | |
| Interview with Heather Piwowar | 92 | | |
| References and relevant links | 96 | | |
| About Heather Piwowar | 97 | | |
| A look into Figshare | 99 | | |
| About Figshare | 100 | | |
| Figshare: Business model | 101 | | |
| Interview with Mark Hahnel | 102 | | |
| References and relevant links | 107 | | |
| About Mark Hahnel | 108 | | |
| A look into Zenodo | 110 | | |
| About Zenodo | 111 | | |
| Zenodo: Business model | 112 | | |
| Interview with Tim Smith | 113 | | |
| References and relevant links | 119 | | |
| About Tim Smith | 120 | | |
| Current state and approaches for further exploration | 122 | | |
| 1. Approaches to reach a critical mass | 122 | | |
| 2. Approaches to change the licensing culture | 123 | | |
| 3. Approaches to avoid data and vendor lock-in | 124 | | |
| 4. Approaches for incentivising innovation | 124 | | |

Preface

Knowledge Exchange (KE) is a collaboration between six key national organisations within Europe tasked with developing infrastructure and services to enable the use of digital technologies to improve higher education and research: **CSC** (csc.fi) in Finland, **CNRS** (cnrs.fr/en) in France, **DAFSHE** (bit.ly/2L488i0) in Denmark, **DFG** (dfg.de/en) in Germany, **Jisc** (jisc.ac.uk) in the UK and **SURF** (surf.nl/en) in the Netherlands. The six partners share a clear vision that scholarship should be open. The partners work together to support the development of digital infrastructures to enable Open Scholarship.

In 2017 KE developed the KE Open Scholarship framework (see page 7 in the report '**Moving from ambition to reality: An account of the Knowledge Exchange workshop on Open Scholarship, Paris 2017, & a menu for possible actions for 2018**' (bit.ly/2BR00h2)) to characterise or categorise activities, discussion topics, proposals, initiatives and more. The framework consists of three dimensions:

- ▶ **Level of granularity (scale)**
 - › macro, covering the population
 - › meso, covering communities/organisations
 - › micro, covering individuals
- ▶ **Research phase**
 - › discovery
 - › planning
 - › project
 - › dissemination
- ▶ **Arena**
 - › political
 - › economic
 - › social
 - › technical

The activity underlying this report targeted the economic arena of the KE Open Scholarship framework. It aimed to collect, describe and learn from examples and initiatives that pioneered change towards new or alternative business models and

payment structures that facilitate Open Scholarship and support it to achieve its full potential. Interviews have been conducted with ten selected initiatives (for more information on the methodology and the interview questions see appendix 1). The result is this collection of interviews that demonstrate successes and challenges to inform and inspire the transition to Open Scholarship.

The report is divided into an **introduction** explaining the background of this activity, followed by **shared challenges** that the interviewees were facing and the ten interviews. It concludes with the **current state and approaches for further exploration**. Moreover, there are some details about the **methodology in appendix 1** and a **glossary in appendix 2**.

Introduction

In the online age, openness and sharing have become, if not the standard, at least household practices in research dissemination. The internet has made it possible to communicate, duplicate, disseminate and reuse scientific digital research outputs using relatively simple means and at little to no extra cost. But the most dominant business models in research dissemination are still only a literal transposition of traditional, printing press-based models and workflows into an online environment. While the value proposition in these traditional models is based on selling or regulating access to resources, this collection of interviews collects a number of examples of organisations and businesses that work along different patterns, with business models based on different value propositions. Every initiative, whether non-profit or commercial, has been selected because they identified a gap in the Open Scholarship landscape and managed to create value by offering a service or product that can help to bridge this gap (for more information on the methodology see appendix 1). They all pioneered a shift towards new or alternative business models and structures that facilitate and support Open Scholarship at its full potential. How to deliver value to the ‘customers’,

how to establish and maintain relationships with end users, and how to generate sufficient revenue streams for the organisation to remain ‘in business’ – these are the challenges that all of the interviewed organisations and businesses are dealing with.

Commercial activity in the field of Open Scholarship is a sensitive topic and so it is not surprising that some of the interviewees objected to describe their organisation as a business or even to framing their activities in a ‘business model’. The concept of a business model has been used to describe the initiative’s value proposition, the customer relations, marketing activities and revenue streams – regardless of whether the initiatives have a not-for-profit or a commercial nature. Within the scope of this report, ‘business’ also encompasses the rationale behind certain decisions about activity range and partnerships but also about intellectual property, collaboration with partners and marketing decisions. As one interviewee puts it: “In the end, whether you call it a business model or something else, it is still an essential part of running your organisation, even if you are a charity”. Every interview includes an infographic that is partially based on the **Business**

Model Canvas as created by Strategyzer (strategyzer.com/canvas/business-model-canvas). Its model is a very useful way to describe any organisation type because of its flexibility to focus on elements other than profit generation. In this collection of interviews, it has certainly helped to analyse the initiatives involved.

These include:

1. **Open Library of Humanities (OLH) (UK): Martin Paul Eve**
2. **OpenEdition (FR): Pierre Mounier**
3. **Opasnet (FI): Jouni Tuomisto**
4. **ASAPBio (US): Jessica Polka**
5. **ScienceOpen (GER): Stephanie Dawson**
6. **HRČAK (HR): Jadranka Stojanovski**
7. **Helsinki University Press (HUP) (FI): Leena Kaakinen**
8. **Impactstory (US): Heather Piwowar**
9. **Figshare (UK): Mark Hahnel**
10. **Zenodo (CH): Tim Smith**

Shared challenges

Although not everything mentioned below is applicable to all organisations, a number of shared challenges and issues faced by most interviewees can be identified. These are related to:

1. **HR and staffing**
2. **Non-profit or for-profit status**
3. **Infrastructure decisions (inhouse development or outsourcing)**
4. **(Open) licensing**
5. **Sustainability and scalability**
6. **Marketing**
7. **Influence on research workflows and the overall position in the research landscape**

Each of these shared challenges will be described in detail below.

1. HR: researchers 'in business'

Almost all interviewees reported staffing challenges: either there is no budget to hire all desired profiles or, if budget is available, interviewees said they are having difficulties finding employees with the right skill set. Of course, financial motives play their part but there is also a link to the typical challenge many

start-ups face when the founder(s) must switch from being the universal fixer(s) to letting other (new) staff members handle certain aspects of the organisation. These one or two person organisations are very much associated with the personal charisma and technical skills of the founder(s) so, in the public perception, the organisation remains very much linked to these personalities. Staffing issues aside, many interviewees mentioned the value of their business/organisation being led by a researcher. Having 'walked the walk' as a researcher is something that the interviewees for whom this is applicable identify as being a determining factor for their success. Once the initiative reaches a certain scale, however, this appears to be a strain for those involved, sometimes even with negative effects on the business operations.

External grant funding turns out to be a crucial element in solving these staffing issues (at least temporarily). Even when they're only a secondary revenue stream, projects and grants provide breathing space and give the initiative the chance to give attention to improvements that help keep their business going, and to innovate and expand their

service portfolio. Therefore, regardless of their business model, almost all interviewees actively pursue grants, even if they consider them as a non-essential source of revenue.

2. Commercial versus non-profit: walking the line

One of the most striking impressions from this collection of interviews is that a ‘hard’ division between commercial companies and not-for-profits when it comes to business models is less strict than imagined beforehand.

When it comes to flexibility, usability and agility – the commercial companies seem to have an advantage. Apparently, as long as there is no data or vendor lock-in, the innovative spirit of (start-up) businesses in the Open Scholarship market is admired (even with some jealousy). However, the commercial initiatives have, for their part, indicated that they feel they struggle when applying for grant funding or to collaborate in funded research projects, either because their legal status doesn’t allow it or, more frequently, because they felt that their commercial status is viewed with distrust by potential partners.

These pangs of regret did not seem to dominate, quite the contrary: a level of frustration with the perceived slowness and inefficiency of non-profits was mentioned by all commercial companies (and some of the non-profits!) interviewed. Several interviewees pointed out that some non-commercial entities (certain university presses and government agencies were mentioned as examples) have certainly not done all they could to accelerate Open Scholarship.

All interviewees stressed that monetising intellectual property of the materials (or of the data collected) is not (and will never be) a part of their business model. Instead, revenue streams flow from the sale of premium services and from licensing tailor-made infrastructures – in addition to a range of services, tools and code available for free. The fear that a successful start-up will inevitably be bought by one of the big commercial players, thus rendering previous guarantees (eg, not to monetise intellectual property of materials) void, has been voiced by almost all non-commercial actors, and is vehemently denied by all commercial companies interviewed. Regardless of whether these claims will hold in the future, a common conclusion seems to be that,

while a non-commercial status does not necessarily mean a better use of available funding, it does offer more protection against abuse such as data lock-in.

3. Infrastructure: make or buy?

None of the initiatives operates entirely on a stand-alone and homegrown infrastructure. In some cases, a host institution provides all or part of the infrastructure. This can include all the physical and organisational structures and facilities needed to run the business or organisation. It covers not only the technical infrastructure such as server space, but also staff-related aspects such as desks and qualified personnel. In other cases, the organisations have usually looked to commercial providers for services related to infrastructure, because it would not be feasible technically and financially to develop an in-house solution. Some interviewees have mentioned that the fear of data or vendor lock-in from external suppliers also plays a role when deciding to develop an in-house solution. Most cases show a pragmatic mix of internal development and externally bought assets.

For most initiatives the situation is not static. Some show an evolution from externally bought infrastructure towards in-house development (this is where the grant funding comes in). Others aim for a mix of sources according to growth and different requirements for their various activities. The interviewees show little hesitation to defend the use of commercial providers even if they themselves are non-profit, although all mentioned that they would never agree to data or vendor lock-in from their suppliers' side.

4. Licensing: pragmatism

When it comes to intellectual property, most of the interviewees showed equal pragmatism. For their own outputs, there is a strong preference for liberal Creative Commons (CC) licences, such as CC BY, and a general aversion to the use of non-commercial licences. Apart from Figshare, who only offer CC BY and CC0 in their free version, all other interviewees refrain from enforcing (if at all possible) these liberal licences too strictly when aggregating or publishing external content. The reasons for this are diverse: they don't have the rights in the first place, or clearing the rights is not feasible, or they are reluctant to impose licensing requirements that the

community doesn't want or need. As mentioned above, non-commercial clauses for own outputs are decidedly unpopular for a range of reasons: either there's no objection against commercial reuse, or the organisation has experienced first-hand that it crucially limits desired forms of reuse. The most common reason against non-commercial clauses was, however, that they are de facto not enforceable.

As mentioned above, in most cases the codebase for infrastructure is open source. In the cases where it's not, the interviewees were quite open about their reasons and ensured that they have built-in precautions against data or vendor lock-in. Nevertheless, some scepticism about the sustainability of these claims in case of a takeover or buy-in has been voiced by some interviewees.

None of the interviewees who had an opinion on the subject defended non-commercial licences (such as CC BY-NC) for their own outputs or codebase. Superficially, this might seem like an obvious way to 'protect' certain assets such as codebase against scooping by the competition, but this didn't turn out to be a viable solution for any of the interviewees.

This does not necessarily mean that they'd encourage competition to scoop their codebase, but enforcing a non-commercial licence is, as one interviewee puts it, "too much of a hassle for it to be efficient".

5. Sustainability and scalability

Although all interviewees want to play a role in a full and mature Open Scholarship market and would like to see their services, tools or the workflows they stand for applied as universally as possible, none of them seems to actively seek monopolisation of their market segment. In other words, an implicit limit to scalability is presumed. For example, the Open Library of Humanities would like to see its model exported to all scholarly literature in the humanities and social sciences (HSS). But, rather than monopolising the market, they'd like to see their library subsidy model applied by more organisations. Figshare is not interested in hosting content licensed other than CC BY or CC0. ASAPBio is focusing solely on biological research. And HRČAK would like to see its model exported to other research outputs at national level but does not aim to become active in other countries.

When it comes to sustainability, almost all those interviewed wanted to move away to some extent from the grant funding cycle and to increase the proportion of their earned income through the sale of services or products. Facing the challenge of trying to generate income from services or products built with commodities that are freely accessible, diversification is seen as essential in order to become sustainable.

As this collection of interviews shows, one of the core challenges faced when the source materials are openly licensed and available for free is generating some form of revenue based on it. It has been mentioned by a couple of interviewees that it would be desirable at international or funder level to create a fund that could be used by developers of open source tools or services as a 'reward' for their effort. Projects and individuals making use of these open source tools or services could allocate a small part of their project funding to such a 'reward' fund.

6. Marketing: build it and they will come?

A good idea and a sound plan are not sufficient to

keep a business going long enough for it to have a significant impact on the Open Scholarship landscape.

There are two clearly identifiable approaches towards marketing amongst our interviewees. One group has never paid much attention to marketing, let alone dedicated staff time to it, but its members are now changing their opinions because they are seeing some clear benefits in a more systematic approach. The others have taken marketing very seriously from the start. This division crosses the commercial/non-commercial divide and seems to be more related to a difference between mature organisations and start-ups. A very typical remark from the first group is that, if they had known that a professional approach to marketing could make such a difference, they would have put more effort in it from the start. For the other group, marketing and communications are a quintessential part of their activities with well-designed campaigns and calculations put in place from the start. At the same time, an example given by Impactstory shows the relativity of marketing efforts: a research profile product they developed gained a lot of traction thanks to the social media savviness of the founders,

but in the end the number of subscribers remained too low for the idea to survive.

7. Influence on workflows and the research landscape

The interviewees were asked explicitly if they felt they are making 'a difference' – a lasting influence on research workflows. The general tone in most interviews is optimistic, not least because of the increasing impact of Open Scholarship in general. There can be no doubt that some of the businesses and organisations have made a lasting impact on the scholarly ecosystem with the most obvious examples being Figshare and Zenodo with their data repositories and the Impactstory algorithms being used by the biggest commercial players in the market. However, the data-related initiatives (Figshare and Zenodo) mention that they think many researchers are using their services because they are a convenient solution for a frequent funder requirement (research data management), not necessarily because they want to make their data open out of principle.



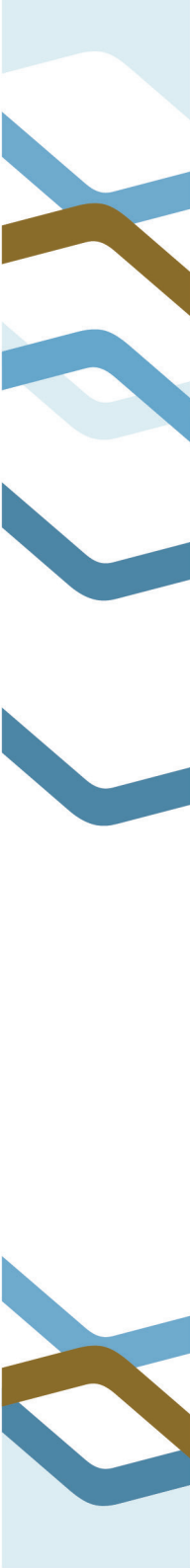
Insights into the Economy of Open Scholarship: A look into the Open Library of Humanities with Martin Paul Eve, Co-founder



About Open Library of Humanities

OLH is a UK charity, with a board of trustees who are responsible for governance. Income generated cannot be used for any purpose other than OLH's charitable goals. OLH is funded via a 'library partnership subsidy model', with over 200 libraries paying an annual sum in order for OLH to publish 24 open access journals that don't require authors to pay article processing charges (APCs), including a mega journal called 'Open Library of Humanities'. Occasionally, OLH also pursues grants to fund tangential activities such as marketing and innovation research.

openlibhums.org



OLH: Business model

Key activities

- ▶ Open access publisher for humanities and social sciences (HSS)



Organisation type

- ▶ Charity
- ▶ Staff: 6 FTE
'(Full-time equivalent)'



Key partners

- ▶ Funding agencies
- ▶ Libraries
- ▶ University presses
- ▶ Infrastructure: Ubiquity Press



Revenue streams

- ▶ Annual membership fees paid by libraries (library consortium model)
- ▶ Grant funding



IP/Copyright

- ▶ Full licence suite for articles published
- ▶ Code base of the in-house platform Janeway: open source



Customers/users

- ▶ Researchers
- ▶ Libraries



Interview with Martin Paul Eve

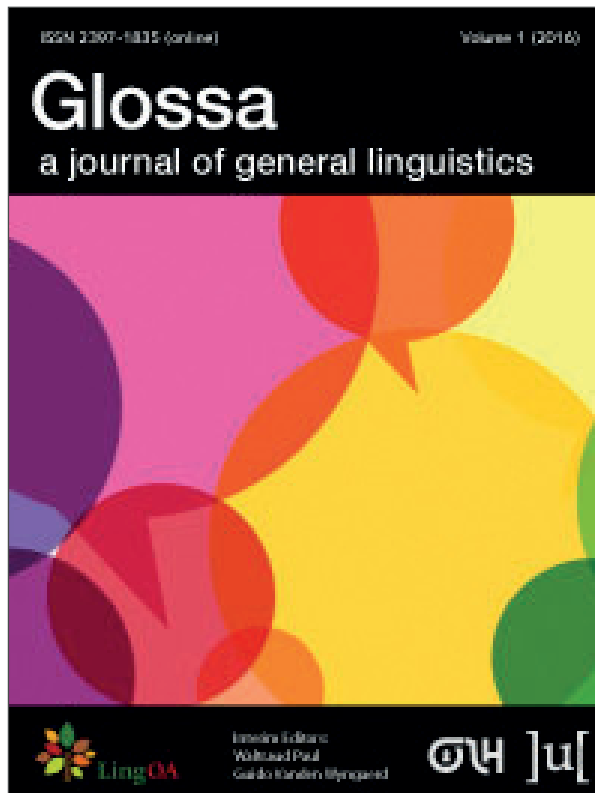
To address the aversion to author fees, considered by most humanities scholars as an unworkable model, OLH developed a new model for open access publishing: the library partnership subsidy model.

“When Caroline Edwards and I established OLH (in 2013) there was a discussion going on in humanities about the applicability of science, technology, engineering and mathematics (STEM) open access paradigms to humanities and social sciences (HSS). The main, successful model for open access publishing model was that of **PLOS** (plos.org); transdisciplinary, author fee-based mega journals seemed the way forward. However, many HSS scholars did not feel that this model was suitable for them and perceived open access as yet another unattainable funder requirement,” says **Martin Paul Eve**.

“Our business model is based on the ‘library partnership subsidy’ model. Our income comes from small annual fees paid by libraries, which makes it look from the outside a bit like the traditional subscription model. Where we are different from the traditional economic models, under which I count the APC or author fee-based open access publishing model as well, is that libraries pay so that we can exist but they don’t necessarily pay for their own benefit.

“All subscriptions are pooled, and subscribing to OLH does not mean that ‘your’ authors can publish or read your publications – because those aspects are already open for everyone. We don’t check if authors are affiliated with institutions that pay a fee to us. We want to say to libraries: ‘you are paying us to be able to exist, and we will publish anyone who passes our peer review process’.

“ Libraries pay so that we can exist, but they don’t necessarily pay for their own benefit. All subscriptions are pooled, and subscribing to OLH does not mean that ‘your’ authors can publish or read your publications. ”



GLOSSA, one of the most well known journals hosted by OLH

“Author fees are not the only issue that needs to be addressed in HSS open access publishing,” says Eve, “but we feared that in coming up with even more innovations, we would lose the community entirely. So, although we do run one mega journal, we decided to remain rather conventional on other fronts, such as using a double blind peer review system.”

Eve sees some real advantages in comparison with author fee-based models: “Author fee-based open access publishers will always have to make sure they reach a certain number of publications to remain sustainable – which in some cases might have an influence on quality control processes.

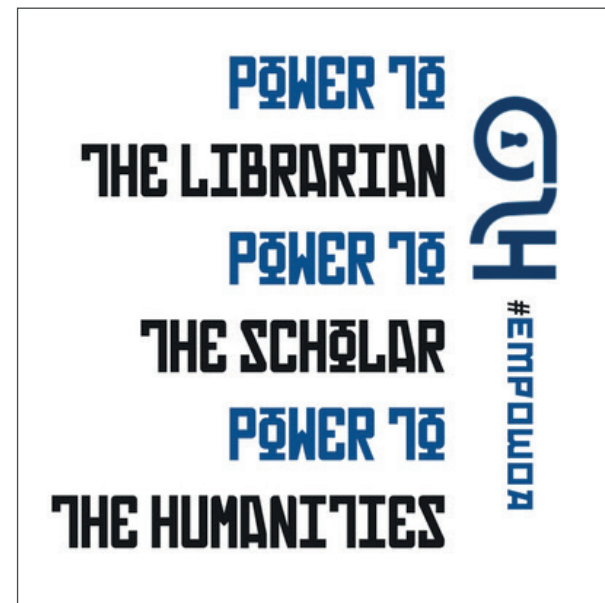
“In our model, we incur a small cost when we publish an article, so we have to ensure good quality control. We keep the fees deliberately small so that individual libraries can circumvent the chain of command related to acquisitions and make the decision to join OLH themselves.”

Currently around 220 libraries are paying an annual fee, which allows OLH to break even. The average production cost per article is around 600 EUR. That includes not only OLH services, but also a buffer to fix mistakes. Eve: “The absolute raw cost per article is probably only around 400 EUR, but overhead costs can increase this quickly, so 600-650 EUR is a safer guess. As we receive the money in advance, via the subscription system, there’s no conflict of interest between our revenue stream and the editorial process. We believe this has a positive effect on academics’ perceptions of our initiatives.”

The OLH business model is dependent on available library funding, supplied with occasional grant funding. Eve: “We currently have a library sign up once every two weeks and a cancellation once every four weeks, so we are growing but at a very moderate pace. One of our current focuses is therefore marketing and outreach, to increase this rate of growth. We have sufficient reserves to keep us in operation for one year. One challenge is to keep institutions renewing their memberships even if we are no longer the new exciting thing. Especially in economic hard times libraries have to make the choice every year between us and Elsevier, because even if they don’t subscribe, they know they’ll keep having access to our content in any case. This is our very own prisoners’ dilemma.”

“ Economic hard times libraries have to make the choice every year between us and Elsevier, because even if they don’t subscribe they know they’ll ‘have’ us anyway. This is our very own prisoners’ dilemma. ”

As library budgets worldwide are decreasing, Eve is worried about OLH’s scalability: “We work with such small margins to keep it cheap, but that gives us very little breathing space. Scaling is a challenge. We receive many more applications to publish journals than we can fund. Even if we increase our fees only marginally in absolute terms, they are so small that we would quickly arrive at a 20% increase, and that worries libraries. They need to give us licence to expand. Libraries vote for journals to come in, but then don’t renew their subscription. They don’t understand that their vote is part of our financial planning. Running a large-scale governance organisation is something that we were a bit too optimistic about. I really underestimated the time and effort these outreach and marketing activities take.



#empowoa: recent marketing campaign by OLH

“If I had known this, I would have tripled efforts in that department from the start. We need to increase our membership rate substantially in order to remain sustainable, but this requires a lot of manpower. It’s not a matter of ‘build it and they will come’ – in most cases, we need to convince libraries in person to sign on and to stay on board. That’s a lot of effort.”

“ We need to increase our membership rate substantially to remain sustainable, but this requires a lot of manpower. It’s not a matter of ‘build it and they will come’.”

Another issue OLH faces is that usage and reader statistics, a factor of interest for libraries when renewing their subscriptions, remain very difficult to gather for open access publications. Eve: “We quite often get requests from big universities for usage statistics when they have to decide to renew their membership, which means we have to spend a long time filling out forms and providing detailed information.

“A lot of staff time is spent on these calculations, which means our operations become more expensive. Also, if you know that the largest fees are only around 2,000 USD, this means a lot of work for us for very little revenue. If everybody requested these detailed statistics, we would have to double our prices. I think libraries often don’t understand the bureaucratic burden for small organisations like ours.”

OLH is a small organisation, with a staff of around six FTE, officially affiliated with **Birkbeck University (bbk.ac.uk)**, where OLH is based. The university deals with technological support, marketing and scholarly communications. Eve: “We are very hands-on and, from the start, Caroline Edwards and I had the best idea of what we were supposed to be doing, including coding the technological infrastructure. The plan for the future, however, is to leave the day-to-day workings to dedicated staff and for us to supervise the academic aspects.”

For its infrastructure, OLH collaborates with **Ubiquity Press (ubiquitypress.com)**: “We started out with them, not only because they offered a good price and were willing to defer author fees during the start-up phase, but also they were very keen to support us.

“We realised we needed a platform to show if we wanted to raise subscriptions, but we did not want to reinvent the wheel by developing it in-house. We are aiming to diversify though, because we do not want to rely entirely on a single partner for our infrastructure. That’s why we decided to build an in-house platform after all, called **Janeway (openlibhums.org/site/janeway)**. A part of our infrastructure is also hosted at the university presses we work with. We pay them to host some journals we make in partnership with them,” says Eve.

“But these presses rely on our support – which is again dependent on our growth – and in any case it remains cheaper for us to publish in-house than to outsource it to a university press. Ideally, we’d find one or more university presses and convince them to implement our model. Most of them are struggling in any case, why not go open access according to our library consortium model and create a big ruckus? For that, I’d have to find the right person, somebody who is not only convinced of the model but also has the institutional power to do it.”

“Most university presses are struggling in any case, why not go open access according to our model and create a big ruckus?”

When it comes to licensing, OLH’s preference is **Creative Commons (creativecommons.org)** Attribution (CC BY) by default, but because there is so much debate about the use of this licence in HSS they do allow more restrictive licences. There

is also a practical reason: because OLH’s portfolio contains a lot of third party content for which reuse rights are not clear at all (a common problem when publishing about art history, for example), the publisher tends to be on the cautious side to avoid lawsuits.

“I am not a licensing purist, I’m a pragmatist,” says Eve. “I prefer ‘free access’ over no access at all. In humanities, a lot of images are being reused and many galleries, libraries, archives and museums (GLAMs) have very unclear licensing policies themselves, so we need to be careful – there’s always the possibility of a lawsuit if we get anything wrong and we cannot afford that. So anything we can get on top of free access is a bonus; if there are restrictions such as only allowing non-commercial use this is not a breaking point for us.”

“I am not a licensing purist, I’m a pragmatist. I prefer ‘free access’ over no access at all.”

“You have to take into account that very few authors understand what they are signing up for when agreeing to an open licence. Some authors even write to me to get my permission to reprint their own, CC BY-licensed works. There is a huge cultural and ethical challenge caused by getting authors to sign something they don’t understand, and publishers should have an obligation to help them understand these terms.”

“Publishers for years have relied on academics not reading contracts, but if we want to claim that we are better than that we have to make an effort to explain to authors the what and why in order to get informed consent.”

“ Publishers for years have relied on academics not reading contracts, but if we want to claim that we are better than that we have to make an effort to explain to authors the what and why in order to get informed consent. ”

OLH is outspokenly not-for-profit. Eve: “Scholarly communications is not a market, and it doesn’t work according to traditional economic rules. You can’t say: ‘I’ll buy this article or that book instead of the one you really wanted’. So why and where do we think that price competition is going to emerge, let alone have a beneficial effect? This is why I’m against commercial actors working in

scholarly publishing, especially the ‘big four’ who are essentially vultures, generating huge profit margins using public funds while at the same time academic libraries and university presses suffer and see their funds being decreased all the time.”

“I think only non-profit actors should be allowed to operate in the field of scholarly publishing. At the same time, some non-commercial actors act as badly as the for-profit ones so that’s no guarantee, but at least it would be a step forward. These giant entities do not spur innovation, as they claim, rather they create monopolies and suck the life out of the small players by buying them up. By taking such a big slice of budgets they inhibit innovation. We have to think about the money you save when cancelling subscriptions, and invest the saved money into innovative infrastructure processes. What’s the role of the library in the 21st century if it is not facilitating open scholarship? Part of the role of the library has to be to enable new and innovative approaches.”

“ We have to think about the money you save when cancelling subscriptions, and invest the saved money into innovative infrastructure processes. What’s the role of the library in the 21st century if it is not facilitating open scholarship? Part of the role of the library has to be to enable new approaches. ”

This outspoken stance doesn't mean that Eve is not interested in the economic aspects of open scholarship: "In the end, whether you call it revenue streams or something else, it's still an essential part of running your organisation even if you are a charity. I don't see why publishing, which is a proper job, should be voluntary or underpaid work. Moreover, if you only work with volunteers, it's not good for diversity because only certain types of people can afford to work without getting paid."

Pioneering the collective library subsidy model, OLH has built a highly respected not-for-profit with an innovative business model. "I would really like to see more experiments with business models outside of the dominant APC model. It's not good practice to hardwire one type of business model into funding requirements. If you invest in different models you can actually generate some market pressure and reduce the risk of monopolisation."

“ I would really like to see more experiments with business models outside of the dominant APC model. It's not good practice to hardwire one type of business model into funding requirements. ”

"We've shown that you can think differently about economics in university environments," concludes Eve. "We hope that we can serve as an inspiration for other organisations to adopt our business model and get a range of publishers doing this, so that we can spread the risk and libraries come to accept that this is a commonplace system. I'd like it if OLH wasn't the only one working this way any more. We need to expand our model and 'normalise' what we are doing, so that libraries will better understand and we don't have to explain it from scratch every time."

References and relevant links

- ▶ OLH: openlibhums.org
- ▶ PLOS: plos.org
- ▶ Birkbeck, University of London: bbk.ac.uk
- ▶ Ubiquity Press: ubiquitypress.com
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ Janeway: openlibhums.org/site/janeway

About Martin Paul Eve

Co-founder



Martin Paul Eve is professor of literature, technology and publishing at Birkbeck, University of London. Previously he was a senior lecturer at Birkbeck, a lecturer in English at the University of Lincoln, UK, and an associate tutor/lecturer at the University of Sussex, where he completed his PhD.

© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A look into the Open Library of Humanities with Martin Paul Eve, Co-founder

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: September 13, 2018

DOI: [10.5281/zenodo.2586862](https://doi.org/10.5281/zenodo.2586862)

This report is available under a [CC BY 4.0](https://creativecommons.org/licenses/by/4.0) licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info



Insights into the Economy of Open Scholarship: A look into OpenEdition with Pierre Mounier, deputy director



About OpenEdition

OpenEdition is a comprehensive open scholarly communication infrastructure for the humanities and social sciences. The OpenEdition portal includes four publishing and information platforms in the humanities and social sciences: **OpenEdition Journals** ([ji.sc/2E5e6v4](https://www.journals.openedition.org/)), **OpenEdition Books** ([books.openedition.org](https://www.openedition.org/)), **Hypotheses** (research blogs) ([hypotheses.org](https://www.hypotheses.org/)), and **Calenda** (announcements of academic events) ([calenda.org](https://www.calenda.org/)). The portal is, thus, a space dedicated to the promotion of research, publishing 700,000 scientific documents that promote open access, while respecting the economic equilibrium of publications.

[openedition.org](https://www.openedition.org/)



OpenEdition: Business model

Key activities

- ▶ Journal platform
- ▶ Book platform
- ▶ Blog platform
- ▶ Event dissemination platform



Organisation type

- ▶ National research infrastructure
- ▶ 50 full-time equivalent (FTE) staff
- ▶ Hosted by four institutions (CNRS, Aix-Marseille University, EHESS, Avignon University)



Key partners

- ▶ Host institutions
- ▶ Ministry for Research and Innovation
- ▶ Publishers and journal editors
- ▶ Libraries



Revenue streams

- ▶ Host institutions
- ▶ French government
- ▶ Project funding (national, regional and European)
- ▶ Freemium services



IP/Copyright

- ▶ All articles aggregated are open access, ranging from full copyright to open licences
- ▶ Codebase open source, licensed general public licence (GPL)



Customers/users

- ▶ Researchers
- ▶ Publishers and journal editors
- ▶ Libraries
- ▶ Government



Interview with Pierre Mounier

OpenEdition started in 1999 as Revues.org, an online platform for two journals in humanities and social sciences (HSS). It soon attracted attention from other journals wanting to join. It was clear that the platform should be open access to increase visibility and accessibility. In 2017, the platform changed its name to OpenEdition. Next to other services, the platform now hosts over 500 journals.

“Today, OpenEdition is a recognised research infrastructure in France. It is even included in the **National Strategy Roadmap for research infrastructures** ([ji.sc/2vWPcZQ](https://www.ji.sc/2vWPcZQ)). We are a “Unité de Service et de Recherche (USR)”, a typical French research infrastructure where efforts are joined to provide a service/platform for the entire research community”, says **Pierre Mounier**, deputy director at OpenEdition.

“Although OpenEdition started out as a platform for journals, we have added other platforms: in 2000 Calenda, a platform for dissemination of HSS events, and in 2009 Hypotheses, a blogging platform. Finally, we moved into monographs and books when OpenEdition Books was added. So now we have a complete infrastructure with four platforms to support scholarly communications in HSS.”

OpenEdition is supported financially by the four founding institutions (CNRS [French National Centre for Scientific Research], Aix-Marseille University, EHESS [School for Advanced Studies in the Social Sciences], Avignon University), which provide the platform with staff, infrastructure and funds to cover operating costs.

They also receive support directly from the Ministry of Research (as a research infrastructure). About 50 FTE staff are permanently seconded from the four founding institutions. The staff are divided into an editorial department that manages the relationships with the content producers (blogging researchers, publishers, journal editors), an IT department that runs systems and development, a department for international development and a department dedicated to the Freemium services - ‘Freemium’ being a pricing strategy by which a digital product or service is provided free of charge, but money is charged for additional features. The other main source of revenue stems from project funding - national, regional and European. These funds are used to develop new and innovative tools and services. Recently, OpenEdition has added the **Freemium model** ([ji.sc/2Vxjge3](https://www.ji.sc/2Vxjge3)) to their revenue streams, but this system has not been introduced to cover operating costs or infrastructures. Rather, it serves to help the journal publishers and editors to cover their publishing and editing costs. Two-thirds of the money collected is transferred to the publishers OpenEdition works with, while the remaining third is retained to operate the commercial services that sell these Freemium services.



“ Permanent funding makes you run the risk that you’ll lose the connection with your users. If you are not attentive to their needs, you’ll gradually lose your users. This can take a long time and you might not even notice it at first because it is not directly reflected in your income streams. ”

“OpenEdition benefits greatly from its status as a national infrastructure. Many similar initiatives are struggling to find sustainability in the long-term, by which I mean longer than just three or four years”, says Mounier. “I’m always a bit surprised when long-term sustainability is considered for only a few years – I think over 20 years! We have this public, national funding that covers our core functions. But the downside is that permanent funding makes you run the risk that you’ll lose the connection with your users. If you are not attentive to their needs, you’ll gradually lose your users. This can take a long time and you might not even notice it at first because it is not directly reflected in your income streams. Users need changes quickly, so this is something that we really have to keep up with.

“It’s important to note that the platform supports the dissemination of the content, but not the editorial process – the responsibility and cost of which remain on the journals – although, with the Freemium model, OpenEdition tries to compensate them partially for their efforts.”

“OpenEdition does not support the editorial process – the responsibility and cost of which remain on the journals – although, with the Freemium model, OpenEdition tries to compensate them partially for their efforts.”

“When we decided to push for open access that was actually the start of a 20 year struggle with a part of the academic community that is traditionally quite reluctant towards open access, although this attitude has changed in recent years. In the beginning, the main driver for this evolution towards open access was OpenEdition founder Marin Dacos, who very early on saw the benefits of it on two levels. A first consideration was that he really saw the advantages and possibilities of the internet and the potential of open initiatives such as Wikipedia: The essence of the web is to be fed by open content. A second consideration was the principle that publicly funded research should be available to the public”, says Mounier. “This principle is especially relevant in HSS, where the societal impact of research has the potential to be very big.” Dacos was able to convince the first journals to not only have an online presence, but to

make the content openly available as well. “Of course, initially there were a lot of objections. There were financial concerns, because journal subscriptions paid for the (often large amount of) editorial work that comes with publishing in HSS. Therefore, publishers saw a threat for their business model. But in HSS, there’s also a more cultural prejudice amongst some researchers who believe that if something is available for free on the internet, it has no value. This rationale, where value can only be added through monetary transactions, had to be argued against and we’ve always had to prove that this is not a valid argument. For example, we have discussions about wordings: We feel that ‘*gratuit*’ (free) is not a good way to indicate the status of an article. Therefore, we insist on using ‘open access’. Making your work open access does not necessarily mean that there are no costs involved, but these don’t have to be carried by the reader.”

In the Freemium programme, OpenEdition provides other types of access than to the html version of the content (such as epub and pdf), against a subscription fee. OpenEdition also offers a metadata feed that libraries can use to feed their

collections, usage analytics and training for libraries who buy these services. Mounier: “I would like to stress that we do not apply any digital rights management (DRM) restrictions – so any individual is still able to create a pdf file from our html versions, for example. So we don’t sell any pdf files, we sell the access to the pdf file, which is a crucial difference. We sell our Freemium services as a package to approximately 500 libraries in France and worldwide and it is not possible to subscribe to a single premium service separately.”

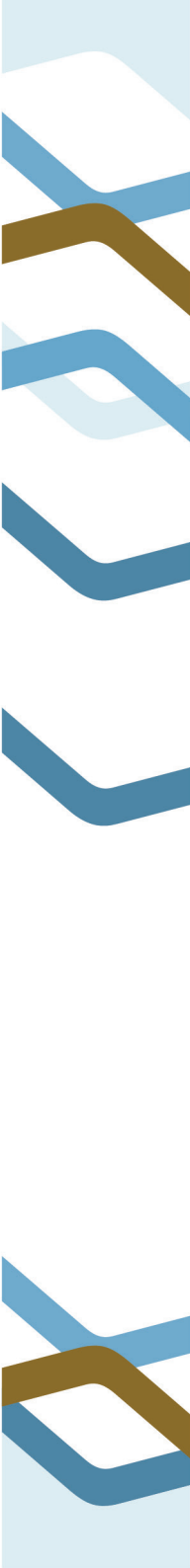
Mounier is not sure that the OpenEdition model can be reproduced in other countries: “We have a significant influence at the national level in France, because we have reached a critical mass – with over 500 journals and 80 publishers and hosting more than 2,000 academic blogs. But I don’t think our model can be copy-pasted to other environments without changes. The research landscape is very fragmented and in every country, and particularly in HSS, structures, habits and stakeholders are different. Topical communities are often very small, and there is always the matter of using local languages. What we are trying to do however, is to help at European

level, with setting up an EU-level infrastructure **OPERAS (operas.hypotheses.org)** – but we don’t try to replicate the OpenEdition model there. For example, we are a centralised infrastructure supported by the national government. In a federalised model, such as in Germany, this would not work. Some countries also don’t have research ministries who are as powerful as in France and in some countries, such as in the UK, the landscape is mainly driven by demand-driven independent Open Scholarship initiatives and services. The French national culture is reflected in the OpenEdition structure, and it cannot be duplicated, I think.”

“ We have a significant influence at the national level in France, because we have reached a critical mass – with over 500 journals and 80 publishers and hosting more than 2,000 academic blogs. But I don’t think our model can be copy-pasted to other environments without changes. ”

Mounier: “In HSS, you have many different types of publishers. Besides private/commercial publishers, you have university presses which are essentially public, you have new publishers which are scholarly-led, and you have university departments and learned societies which are also publishing journals. I believe that private publishers are important to perform certain functions, despite their different approach. I think that private and public publishers can co-exist without difficulty, but when it comes to infrastructure there are issues. Publishing platforms have a central place in the Open Science ecosystem. If they are privately owned this is a problem for the entire system, especially when they apply strong DRM technology and impose it on all users – which limits interoperability – or when they apply ‘lock-in’ strategies that trap users in their services. So I believe that infrastructures and platforms should not be privately owned but community owned and/or publicly funded.”

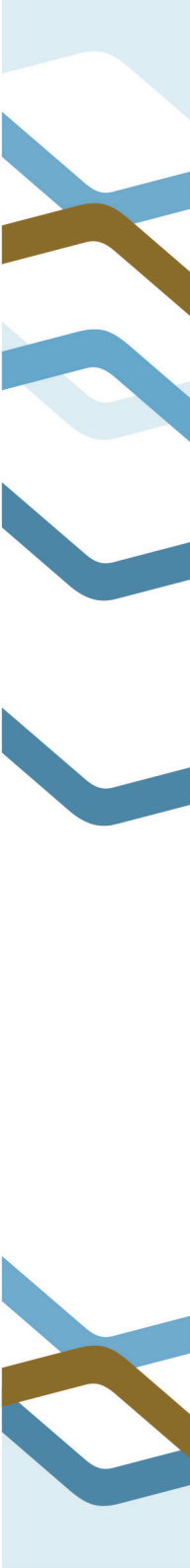
“ Publishing platforms have a central place in the Open Science ecosystem. If they are privately owned this is a problem for the entire system, especially when they apply strong DRM technology and impose it on all users – which limits interoperability – or when they apply ‘lock-in’ strategies that trap users in their services. ”



OpenEdition has developed its own content management system (CMS) open source software **LODEL (lodel.org)**. “In 1999 there were not that many systems that were suitable for HSS online publishing, because of the highly complex content structuring that is needed”, says Mounier. “Now, of course, there are many other options, such as **Open Journal Systems (OJS) (openjournalssystems.com)**, which are very efficient for these purposes. LODEL is openly licensed with **GPL (ji.sc/2Vh6Xxm)**, but we have always struggled to make it really usable outside of OpenEdition. When you develop open source software, it’s not enough to apply the licence and make the code available. You have to package the software, you have to offer community support, etc. We have made some efforts in this area, but to do this right we don’t have the resources available. In the past our position has been that we develop the software for our own needs, we openly licence it and document it as well as possible, but

then the users are mostly on their own. So the usage of our software outside of OpenEdition is rather low, although some French universities have installed their own local platforms based on LODEL. For example, with OJS, it was really part of the Public Knowledge Project’s mission to have their software distributed as widely as possible and to help the community to do this. We like to work with them and learn from their experiences, but this has not been part of our core mission so far. For our blogging platform, we decided that LODEL was not a good solution. We felt that it would be easier to run it on WordPress. This is not without difficulty though – we have no control over the development of this software and versioning, and communicating changes to our users, for example, can be very frustrating, such as when Wordpress introduced a new content editing workflow and everybody was very confused for a while!”

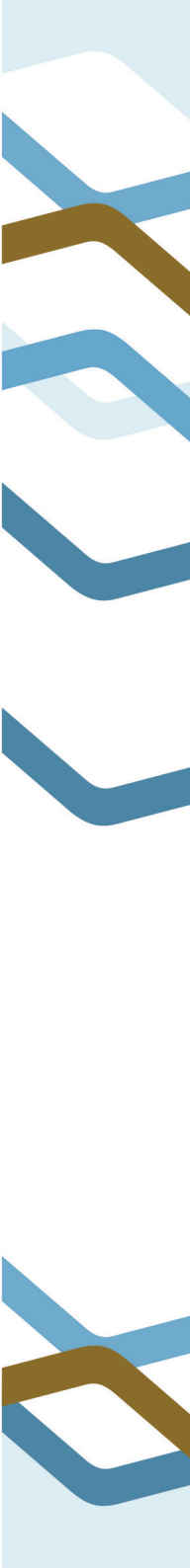
“ Our position has been in the past that we develop the software for our own needs, we openly licence it and document it as well as possible, but then the users are mostly on their own. ”



OpenEdition is in favour of open licences, but it is not always possible to enforce them. Mounier: “In HSS, resistance towards open licences is very strong. Most of our journals still apply a classical copyright regime, although we see a growing adoption of **Creative Commons** (creativecommons.org) licences. However, it’s almost never the most liberal CC BY licences – but rather CC BY-NC and CC BY-NC-ND. We advocate actively for the adoption of liberal licences, but we cannot impose them and many journals, actually, have very good arguments against the most open ones. For example, OpenEdition has come out supporting the option of the ND (no derivatives) clause for **HSS in a reaction to the Plan S Implementation Guide** (oep.hypotheses.org/2169). We are already very happy if we can get them to move from classical

copyright to CC BY-NC-ND. Our best argument is that the latter allows for easy reuse in a non-commercial environment such as teaching and research, which is something that a lot of researchers are likely to be doing. So that can work. The main obstacle is that in many cases, researchers don’t want to bother about it. Classical copyright is easier, they don’t have to discuss with their editorial committees about it. But nonetheless, in our application form we ask about the licences, and it links towards Creative Commons – and we get questions about open licences because of that. Many editors could be interested in the possibilities if it is explained to them properly.”

“ Classical copyright is easier, [researchers] don’t have to discuss with their editorial committees about it. But nonetheless, in our application form we ask about the licences, and it links towards Creative Commons – and we get questions about open licences because of that. ”



“ Open scholarship is an ongoing process that supports the entire research lifecycle, not only the dissemination of end results ”

“Open Scholarship is a very broad concept that can include a lot of different activities,” says Mounier. “We deal with scholarly communication and open access publishing, but I think the most interesting move we made was starting the blogging platform Hypotheses. This opened up the entire research process from only communicating the end results to the progressive opening of the research process itself. It allows us to show what’s happening before publication and it offers an incentive for researchers to ‘open up their workshop’, to communicate during their activities what they’re doing to their colleagues and to the public. Open Scholarship is an ongoing process that supports the entire research lifecycle, not only the dissemination of end results.”

Mounier thinks that the position of HSS in the academic environment and in Open Science is problematic: “In terms of funding and policies, developments in Open Science are really science, technology, engineering and mathematics (STEM) driven. Policy makers are really pushing for Open Science practices and policies, but they have mostly the STEM model in mind. HSS have specificities and for me there is a risk that the STEM model will take over in Open Science and that HSS will be more and more marginalised in the movement. We see it as our mission to represent HSS stakeholders in the Open Scholarship ecosystem. An example is open data, which is a core element of Open Science. But when people talk about open data, they often mean big data and data-crunching and this does not reflect what is happening in HSS. In our communities, publications and monographs are the basis of research and research data in HSS is often small data. We need to ensure that these continue to have their place in Open science.”

References and relevant links

- ▶ OpenEdition main site: openedition.org
- ▶ OpenEdition Journals: [ji.sc/2E5e6v4](https://www.ji.sc/2E5e6v4)
- ▶ OpenEdition Books: books.openedition.org
- ▶ Hypotheses: hypotheses.org
- ▶ Calenda: calenda.org
- ▶ ‘French National Strategy On Research Infrastructures, 2016 Edition’: [ji.sc/2vWPcZQ](https://www.ji.sc/2vWPcZQ)
- ▶ OpenEdition Freemium: [ji.sc/2Vxjge3](https://www.ji.sc/2Vxjge3)
- ▶ OPERAS: operas.hypotheses.org
- ▶ LODEL: lodel.org
- ▶ Open Journals Systems (by Public Knowledge Project): openjournalsystems.com
- ▶ GNU General Public Licence (GPL): [ji.sc/2Vh6Xxm](https://www.ji.sc/2Vh6Xxm)
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ OpenEdition Declaration on the Plan S Implementation Guidance (19th February 2019): oep.hypotheses.org/2169

About Pierre Mounier

Deputy director



Pierre Mounier is deputy director of OpenEdition. He has published several books about the social and political impact of ICT (Les Maîtres du Réseau, les enjeux politiques d'Internet, [The network masters: the political challenges of internet] 2001), digital publishing (L'Édition électronique [Electronic publishing], with Marin Dacos, 2010) and digital humanities (Read/Write Book 2 – Une introduction aux humanités numériques [Read/Write Book 2: An introduction to the digital humanities], 2012; Les Humanités numériques – Une histoire critique [A critical history of the digital humanities], 2018). As deputy director of OpenEdition, Pierre Mounier's work mainly revolves around the development of an internationalisation strategy for the infrastructure, in particular by establishing partnerships with platforms and institutions in Europe and elsewhere. To further this objective, he regularly participates in international conferences and seminars to present OpenEdition's programmes and discuss subjects relating to digital humanities and open access. Pierre Mounier coordinates the development of OPERAS, a European infrastructure dedicated to open scholarly communication in HSS, gathering 39 partners from 15 countries.

© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A look into OpenEdition with Pierre Mounier, deputy director

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: February 28, 2019

DOI: [10.5281/zenodo.3084406](https://doi.org/10.5281/zenodo.3084406)

This report is available under a CC BY 4.0 licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info



Insights into the Economy of Open Scholarship:

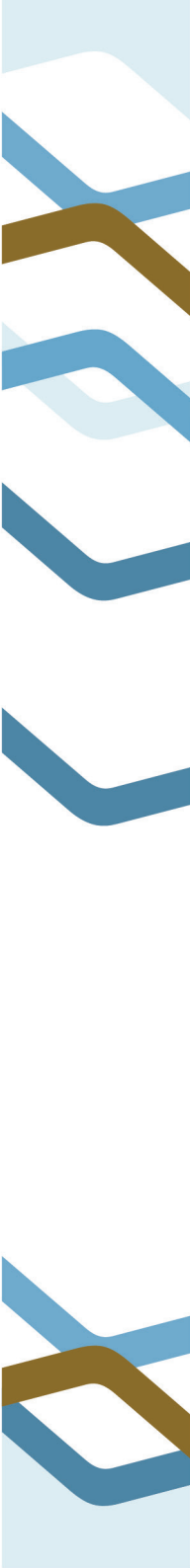
A look into Opasnet with
Jouni Tuomisto, co-founder



About Opasnet

Opasnet is a wiki-based website and workspace for helping social decision-making. Opasnet is maintained and developed by the **National Institute for Health and Welfare (THL)** (thl.fi/en/web/thlfi-en) in Finland. The website collects, synthesises, and distributes scientific information and values. Opasnet is run by a small research group at THL. After a startup phase where the initiative received funding from various projects, enabling the researchers to build the platform, it now receives a small but stable level of funding from THL.

opasnet.org



Opasnet: Business model

Key activities

- ▶ Wiki-based platform
- ▶ Recently launched: concept of knowledge crystals



Organisation type

- ▶ Research group



Key partners

- ▶ National Institute for Health and Welfare (THL)
- ▶ Civil society groups (potentially)



Revenue streams

- ▶ Infrastructure and staff hosted by the National Institute for Health and Welfare (THL)
- ▶ Grant funding (previously)



IP/Copyright

- ▶ All outputs CC BY-SA 3.0



Customers/users

- ▶ Individual researchers
- ▶ Research groups
- ▶ Project partners



Interview with Jouni Tuomisto

Opasnet is a wiki-based platform, which means that it is a website or database developed collaboratively by a community of users that allows any user to add and edit content. It has many similarities with the design of **Wikipedia** (wikipedia.org). While co-founder Jouni Tuomisto rarely produces original data, he synthesises research about environmental health and impact assessment issues and creates models based on that research.

“I started Opasnet in 2006, after I heard about Wikipedia for the first time,” says **Jouni Tuomisto**. “I took some time to look into how it worked, I was really impressed and started to think that this wiki-based approach was the way science should be done. It was so much more efficient than any other system I knew! Because my work synthesises other research, it is a good area for wiki-based work. I can describe the context and analysis online, discuss the content, and then feed it all into the models. I can then show everything online. Such tools did not exist at the time and introducing the wiki format into my line of work was really a revolution.

““ Introducing the wiki format into my line of work was really a revolution. ””

“The design and software we use for Opasnet is based on what Wikimedia does – we also licence all our outputs CC BY-SA 3.0 (**Creative Commons, Attribution-ShareAlike - creativecommons.org**), mimicking the Wikipedia

policy. I have been in low-key collaboration with the Wikimedia community, but my role is not an encyclopaedist so our focus is not the same. The main difference is that Wikipedia collects existing information, whereas in Opasnet we are producing new information. We take research data and studies and try to make policy-relevant syntheses that cannot be found in textbooks. But I’ve learned a lot from their processes and, as I don’t believe in reinventing the wheel, I think their policies are the most solid ones for providing a true open source environment.”

During its startup phase, Tuomisto’s research group was working on three large grant-funded projects. Thus, until 2011, Opasnet was very well resourced. As a result, Tuomisto’s research group was able not only to develop the wiki (using **R software r-project.org**) but also to do meta-research on open workflows and the infrastructure needed for that.

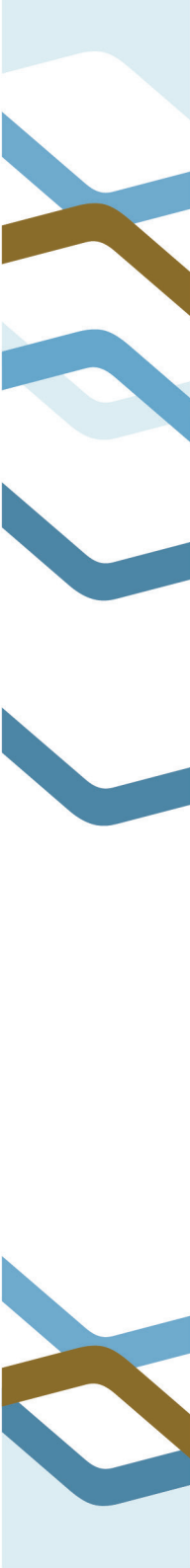
Tuomisto is convinced that this theoretical work about information flows has helped the research group a lot when maintaining the wiki: “A student of mine even wrote a thesis on the way we worked – what kind of structures you need if you are working in an open workspace, what kind of objects you should produce so that they can be criticised and reused. We could also do a lot of practical experimentation with what works and what doesn’t.”

Since 2014, Opasnet has not received much funding, except a small but stable level of support from the THL. The research group has continued to do limited development work on Opasnet and has been able to develop it into a workspace with all the functions the researchers need. The core group of Opasnet users is very small, mainly the seven members of the THL research group that developed it. Occasional, but much less intensive, use has been logged from within THL and also from outside the organisation.

Tuomisto: “We are now being funded enough to maintain and update the system and to keep the website usable, but we haven’t been able to develop it further technically. However, because we had such a head start during our funded period, I believe we got most things right from the beginning – there hasn’t been a huge need for major updates or further development.”

“ Our research group has continued to do limited development work on Opasnet and has been able to develop it into a workspace with all the functions the researchers need [...] if we want to do some more innovative work again, we’ll need more team members working on the wiki. ”

“That doesn’t mean extra funding wouldn’t be welcome – if we want to do some more innovative work again, we’ll need more team members working on the wiki. We have a current page load of 90,000 per year. I would like to see this number increased. I would love to connect more with people outside of THL as well. Recently, we successfully participated in a hackathon, and for me one of the most exciting outcomes was that I got to connect with **Open Knowledge Finland (okf.fi)** and with the **National Library of Finland (kansalliskirjasto.fi/en)**.”



“In general, I believe overhauling the entire research workflow in my field is a very hard thing to do. I have convinced many people to take one or two small steps, but I have not yet managed to get them to work entirely in the open. I have been trying to convince people to work in open workspaces, but it remains a very controversial idea,” says Tuomisto.

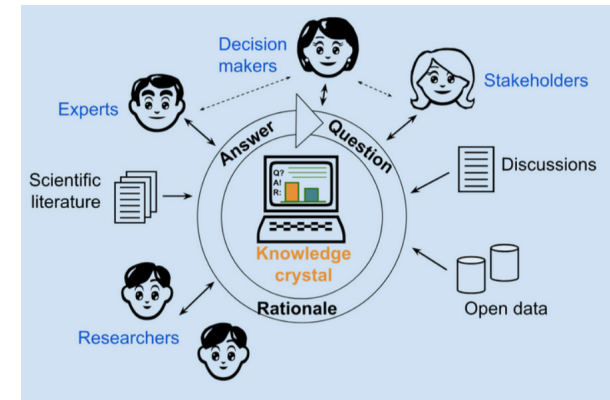
“Most researchers like the idea of wikis in theory, but in practice they remain convinced that they need to publish their research the old way.”

“Most researchers like the idea of wikis in theory, but in practice they remain convinced of the need to publish their research the old way. They only want to open up their data after the publication of an article, and in practice that often doesn’t happen at all – although, quite recently, research data management (RDM) has received some traction because funders are starting to require it. Usually, you spend [all your

funding] developing results, and you only start to think about sharing your data at the end, when the funding has ended – so the incentive to share the data for any given project is small.

“After the project has ended you start a new one and the cycle restarts, and a lot of information just remains inaccessible forever. In my experience, if you don’t open your data at the moment you create it you just never get around to doing so, so I try to convince people to be open from the start but I have not been very successful. In practice, this meant that I used to argue a lot with my colleagues about this. Unfortunately, in project meetings I was often overruled by a majority vote on these matters!”

Quite recently, Opasnet launched the term ‘knowledge crystals’, defined as ‘current best answers to specific research questions, produced and distributed openly using crowdsourcing and scientific criticism’.



Schematic overview of a knowledge crystal

The idea of a knowledge crystal is to combine only the useful parts of information products to support decision-making. An information object is built around a specific research question. The question can be purely scientific but, in the case of decision support, it is usually phrased to precisely address a future decision. To answer the question, experts gather all possible material that could help – mainly research articles, but also research data and expert reports. The strength of a knowledge crystal is that it combines the best of three worlds: it can use all relevant information (not only the researcher’s own data as in a traditional research article), it interprets the data (unlike open data), and it is produced by following the principles of openness and critique (unlike an expert report).

Tuomisto: “As researchers, we have identified three key principles that we’re set to obey at all times – and we’ve given these principles a practical face with the knowledge crystal. The first principle is that all of our work has to be open at all times. The second is that whatever we do must be made available for criticism. The third is that whatever we

produce has to be organised by a topic or research question and stored in a permanent location. Everything is always put in the same place. We improve our answers to the research question but the question can always be found at the same internet address, thus making it possible to develop machine-readable interfaces to the answers.

“These three basic ideas are our guiding principles and we don’t accept any activity that is contrary to them. With our knowledge crystals, we think we have created a tool that will convince others to follow these principles as well – because we have managed to make scientific information clearer and more relatable.”

With the knowledge crystal concept, Opasnet won a hackathon organised by **Helsinki Think Company** (thinkcompany.fi/portfolio/wide) and the National Library of Finland. As a result, the research group has received quite a lot of media attention and collaboration requests.

“Knowledge crystals are an old concept in a new package,” says Tuomisto. “We used to use the terms ‘variables, methods, and assessments’, but these are not very accessible words and are perceived as too technical and complex to be useful. Since we won the hackathon, for the first time I believe we’ve managed to get some real traction around our work. We’ve received media attention and a lot of personal messages from people who are interested in the idea.”

“ I’m really bad at marketing, but the hackathon experience has taught me that it is useful to have an attractive product. ”

“Even within THL, there’s suddenly an interest in what we’re doing! Of course, knowledge about open science concepts has evolved over the last ten years, so maybe people are simply more prepared to listen to us now. In any case, I believe we should try to promote the concept of the knowledge crystals further; from a marketing perspective, it can also be a useful tool to raise funds to continue work on Opasnet. I’m really bad at marketing, but the hackathon experience has taught me that it is useful to have an attractive product. It’s exciting to see others grasping the opportunities, not only for research but also as a basis for policy making.”

*[subsequent to the interview, the city of Helsinki started to use knowledge crystals for the implementation of their **Carbon Neutral Helsinki 2035 Action Plan** (hnh.hel.ninja/), Gwen Franck]*

A wiki-based workspace is a useful tool for research groups to work in a collaborative environment. The question remains whether initiatives like Opasnet are scalable, and whether the same workflows can work for bigger groups of researchers in other fields.

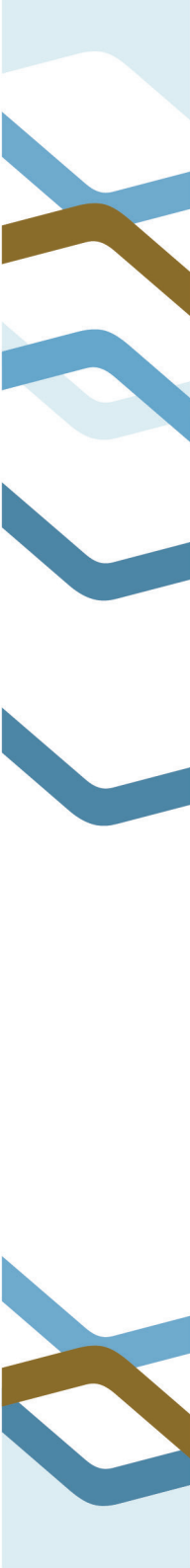
According to Tuomisto, the biggest strength of Opasnet is that it obeys the basic principles as to how science should be done: “Despite our small size, we have been more faithful and more successful in this aspect than most initiatives. We can be a beacon for other researchers, by showing how open science practices are not only better but also more efficient than closed principles. As the most important things are ideas and good practices, supported by open source software, our system is easily scalable. You can join the Opasnet community or copy our code and start your own web workspace, without any restrictions.”

A lot of for-profit initiatives offer similar services to researchers, allowing them to collaborate and share their research. Yet, despite none of them having the flexibility of a wiki, and open science principles being entirely or partially compromised when using these platforms, they remain the most popular solution for most researchers. “Opasnet is not very competitive, despite the effectiveness of our principles,” says Tuomisto. “The initiatives that will get traction are usually the most adaptive and most productive

ones, but these are not necessarily the most open. Commercial platforms are able to develop user-friendly appeal and provide solutions to the immediate needs of researchers, even if their fundamental principles such as intellectual property rights or openness go against the main principles of science. I don’t see myself as an opponent of private companies, who run similar systems for money. I am not fighting against publishers; I simply think their product is not as good as the open solutions. I think open solutions are mostly better and have more impact potential.

“Personally, I have avoided data or vendor lock-in because I am abiding by my open principles, but sometimes colleagues are stuck with their current closed systems and that can make collaboration more difficult.”

“ I am not fighting against publishers; I simply think their product is not as good as the open solutions. ”



Tuomisto is convinced that the shift towards more open research workflows is not, for most researchers, a natural process. “Imposing open science principles at a higher level, such as via funder or institutional policies, is therefore essential to promote the transition towards open scholarship. I think there’s more room for policy initiatives. Most researchers simply want to do science. They don’t see it as their role to fight old-school publishers and don’t want to endanger their careers. They acknowledge that ‘open’ could be a solution, but they won’t fight for it. If, for example, the **Academy of Finland** (aka.fi/en) endorsed open science principles, most researchers would be happy to comply. But they don’t necessarily know how to do that.”

“Imposing open science principles at a higher level, such as via funder or institutional policies, is essential to promote the transition towards open scholarship.”

“Incentives should be designed to promote open science, rather than punishing it. At the same time, however, it’s essential to acknowledge the efforts previously made – and I don’t have a ready-made solution for that. Luckily, many problems don’t arise until you effectively start to work in the open, and they can then be solved gradually – you don’t need to predict all potential issues and provide solutions for them in advance,” concludes Tuomisto.

“You don’t need to predict all potential issues and provide solutions for them in advance.”

References and relevant links

- ▶ THL: thl.fi/web/thlfi-en
- ▶ Opasnet: opasnet.org
- ▶ Wikipedia: wikipedia.org
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ Open Knowledge Finland: fi.okfn.org
- ▶ National Library of Finland: kansalliskirjasto.fi/en
- ▶ Helsinki Think Company: thinkcompany.fi/portfolio/wide
- ▶ Carbon Neutral Helsinki 2035 Action Plan: hnh.hel.ninja
- ▶ Academy of Finland: aka.fi/en

About Jouni Tuomisto

Co-founder



Dr Jouni Tuomisto has a degree in medical sciences, and over 25 years of research experience in environmental health. He is a chief researcher at THL. His research focuses on health impact assessment and decision analysis. The work of his research group is supported by the wiki-based workspace Opasnet. The group has produced a set of recommended practices for decision support and evaluation, and for the management of decision processes.

© Knowledge Exchange 2019

Title: Insights into the Economy of Open
Scholarship: A look into Opasnet with
Jouni Tuomisto, co-founder

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: October 30, 2018

DOI: [10.5281/zenodo.2842456](https://doi.org/10.5281/zenodo.2842456)

This report is available under a CC BY 4.0
licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info



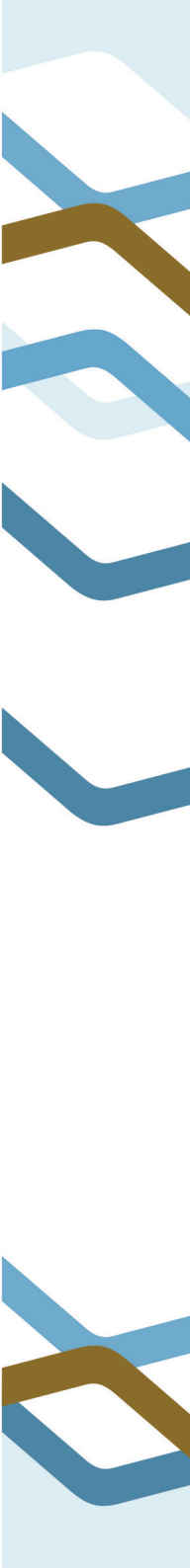
Insights into the Economy of Open Scholarship: A look into ASAPbio with Jessica Polka, Executive director



About ASAPbio

Accelerating Science and Publication in Biology (ASAPbio) is a scientist-driven initiative to promote innovation and transparency in life sciences communication. ASAPbio is a nonprofit incorporated in the state of California. It receives grant funding and also has a member advisory board consisting of six funders who financially contribute to the activities of the organisation.

asapbio.org



ASAPbio: Business model

Key activities

- ▶ Advocacy about preprint sharing
- ▶ Advocacy about open peer review
- ▶ Research and monitoring of research funder policies on preprints



Organisation type

- ▶ Non-profit
- ▶ Staff: 2.25 FTE



Key partners

- ▶ OpenUP project, TRANsparency in Scholarly Publishing for Open Scholarship Evolution (TRANSPOSE)
- ▶ Public Library of Science (PLOS)
- ▶ Advisory board funders
- ▶ Creative Commons



Revenue streams

- ▶ Advisory board of six funders
- ▶ Grant funding



IP/Copyright

- ▶ Advocacy for CC licences on preprints
- ▶ Own outputs: CC BY



Customers/users

- ▶ Funders
- ▶ Researchers



An interview with Jessica Polka

ASAPbio started as a group of four biologists (Jessica Polka, Daniel Colon-Ramos and Harold Varmus, led by Ron Vale) at University of California, San Francisco (UCSF) (ucsf.edu). They came together as members of various life science research groups, with the mission of establishing better and more sustainable research practices in the life sciences.

“Despite all the advantages of the digital tools available today, the speed of the actual communication has not increased and, as a result, science overall suffers,” says **Jessica Polka**, executive director at ASAPbio. “Ron Vale had already written **an article ([pnas.org/content/112/44/13439](https://www.pnas.org/content/112/44/13439))** about this in 2015, which showed that students at UCSF needed more and more time to complete their degrees, because the time and work needed to put together a paper (as first author) and get it published – a necessity to advance an academic career – has increased so much over the years. In this paper, Ron presented the publishing of preprints as a possible solution.”

At that time, two popular preprint servers had emerged: **bioRxiv (biorxiv.org)** and **PeerJ Preprints (peerj.com/preprints)**. At first, the uptake was relatively low, but the group saw the opportunity and they organised a meeting at the beginning of 2016 to try to understand whether preprints could play a bigger role in the life sciences.

Over 70 scientists, publishers, funders, and other stakeholders gathered to talk about the potential benefits preprints could have in accelerating the speed and efficiency of scientific communications. In part because this workshop was so successful, they managed to get grants from four different funders to push the work forward as ASAPbio. ASAPbio is entirely grant based and does not supply any direct services. Polka says that they don't intend to change this, though they would like to diversify their sources of support, for example, through participating in research projects.

ASAPbio started to monitor changes in the environment, such as the potential effect of funder policies that encourage and validate the usage of preprints. Perhaps the most important aspect is the inclusion of preprints in more formal infrastructures, for example, including preprints in **Crossref (crossref.org)** so that digital object identifiers (DOIs) can be issued for them.

ASAPbio's preprint work is supported by funders. Six of them form an advisory board and support the work through contributions. "Preprint publishing has so many benefits for the authors, it removes barriers to openness, such as journal embargoes. Acknowledging that it is a form of work that should be recognised by funders as a proof of activity is an essential part of our activities," says Polka.

"It has been very exciting to see that the involvement of funders in encouraging positive preprint policies is increasing. Allowing them to be citable on grant requests, for example, can be really crucial. The **National Institutes for Health (NIH) (nih.gov)**, one of the biggest funders for life sciences in the US, allowed this, and it has had an enormous effect on legitimising preprints."

“ It has been very exciting to see that the involvement of funders in encouraging positive preprint policies is increasing. ”

Although the idea of treating preprints as a method of science communication in its own right has gained a lot of traction over the last year, especially in the life sciences, not everybody is convinced of its merit. Polka: "The biggest hurdle to accepting preprints as a recognised form of research output is that the fear of scooping is very present."

"There are other arguments against the practice, but I feel that we can more easily counter those. For example, there is the argument that it could lead to quality decline. This is, in my opinion, a false argument because we as researchers are already constantly sharing our unpublished work in conferences and meetings, posters, and talks. This is not necessarily peer reviewed work. There is also a fear that people will share low quality information once they are able to share preprints, but I think people will always be worried about their reputation and they will not be inclined to share low quality work. The issue of scooping is a bigger one, however. If not everyone respects preprints as a legitimate form of scientific communication, a competitor might see it as an opportunity to scoop research. That's why I think the concept of being able to cite them properly is so important. In a way

they should be treated just like regular journal articles, provided that it's clearly indicated that they're preprints."

“ If not everyone respects preprints as a legitimate form of scientific communication, a competitor might see it as an opportunity to scoop research. That's why I think the concept of being able to cite them properly is so important. ”

A much debated topic is the copyright status of preprints. In a recent collaboration with **Creative Commons (creativecommons.org)** and **PLOS (journals.plos.org)**, ASAPbio has created some **resources (asapbio.org/new-licensing-resources)** that deal with preprint licensing in detail. These include an FAQ aimed at researchers, answering questions such as 'does the act of posting a preprint transfer copyright or sign transfer rights away to the preprint server provider?' and 'why should authors consider applying an open licence to their preprints?'

How open is your preprint?

The license you choose has a big impact on how your work will be shared & reused.

The Creative Commons (CC) licenses described here break down the barriers to sharing by communicating rights and permissions up front with everyone.

CC0 waiver
CC0 places work in the public domain, waiving all copyright and related rights. Allows anyone to reuse your preprint in any medium for any purpose, even without attributing it to you. Often used for works created by U.S. government employees, as these are already in the public domain in the U.S. Ideal for datasets.²

CC BY
Attribution (BY) Allows anyone to repost or modify your preprint in any medium for any purpose, but requires that users provide attribution to you and include a link back to the original whenever the material is used and shared. Encouraged by NIH.¹ Fits the original definition of open access.³

-NC, -ND, -SA
These terms can be added to the CC BY license to produce 5 other licenses: <https://creativecommons.org/licenses/>

- Noncommercial (NC)** Prohibits commercial use of the material. If you select it, you don't grant permission to:
 - Republish a figure in a paywalled journal
 - Use the preprint to advertise products
 - Reprint the work in a textbook sold commercially
- No derivatives (ND)** Prohibits the sharing of adaptations of the material. If you select it, you don't grant permission to:
 - Translate the preprint to another language
 - Create a copy of the preprint with extensive annotations
 - Adapt a diagram or drawing for use in another paper
- ShareAlike (SA)** Requires adaptations of the material to be released under the same license. For example, a figure that is modified from your preprint would have to also be published under a CC BY-SA license. (However, a book containing that modified figure could have its own, more restrictive license). This license is used by Wikipedia and Wikimedia Commons.

No license
All rights reserved. If you do not select a license, you do not give default permission to reuse the work (beyond what is required to post to the preprint server). As a result, you don't grant permission to:

- Repost your paper, unchanged, on a class website
- Using a figure in academic talks or text & data mining may also be prohibited in countries without a fair use or equivalent doctrine. Note that some servers (bioRxiv, etc) allow TDM for all manuscripts.

Remember...

- All CC licenses require reusers to indicate if changes have been made, which alerts others that the work as modified is not the same as the original.
- As long as you retain the copyright in your work, you can always grant additional permissions on an individual basis. This includes giving permission for someone to reproduce or modify your work, commercialize your work, or transferring copyright to a journal or signing a license to publish agreement.
- Licenses are permanent, but don't stop authors from releasing other versions under other licenses.
- Professional norms for citation and plagiarism apply regardless of how content is licensed, and even for works dedicated to the public domain under CC0. Often, those norms are more restrictive than the attribution requirements of CC licenses.
- Fair use and other limitations and exceptions apply regardless of which license is selected.

References

- <https://grants.nih.gov/grants/guide/notice-files/NOT-00-17-050.html>
- <http://www.soros.org/openaccess/boai-10-recommendations>
- <https://creativecommons.org/share-your-work/public-domain/cc0/>

CC0: from <https://creativecommons.org/about/downloads>
Open access: from <https://creativecommons.org/wiki/faq/faq-open-access>
See original at <https://creativecommons.org/licenses/by-nc-nd-sa/>
This work is licensed under a Creative Commons Attribution 4.0 International License. <https://creativecommons.org/licenses/by-nc-nd-sa/>

How open is your preprint? Resource created by ASAPbio asapbio.org/new-licensing-resources

Creative Commons and PLOS asapbio.org/new-licensing-resources

Another resource created together with Creative Commons and PLOS was a one-page infographic called ‘How open is your preprint?’, intended to encourage authors to apply the most open licence possible to their preprint.

Polka: “I don’t know of any preprint server that requires authors to transfer their copyright to post. They have to provide at least a basic licence that allows the server to publish the paper, but they can also use a more liberal Creative Commons licence to allow more forms of reuse. Unfortunately, researchers are not always aware of the different licensing options and there’s also a lot of uncertainty about how the final journal version will interact with the preprint. In practice, however, I only know of one publisher that has a policy disallowing CC licences on preprints. In general, when the author retains their rights, they are free to relicence and renegotiate. Publishing a preprint doesn’t necessarily have to undermine the relationship with the eventual publisher.”

“In general, when the author retains their rights, they are free to relicence and renegotiate, so publishing a preprint doesn’t necessarily have to undermine the relationship with their publisher.”

Polka: “I’m personally in favour of using very liberal open licences for preprints and papers. They should not only be free to read, but the user should also be allowed to do other things with the content. Everything on our own website is Creative Commons Attribution (CC BY). But I do use proprietary software and social media myself, so I am making a lot of compromises against these principles myself. For instance, I use Zenodo, but I create my slides with Google, so I might have a double standard in my daily life.”

“Working fully ‘in the open’ is often more complicated than it seems, but I hope that the choices we make as ASAPbio reflect the idea that scientific information should be as open as possible. I believe that, regardless of the licence you choose, knowing the exact ramifications of applying that licence is essential.”

“ I believe that, regardless of the licence you choose, knowing the exact ramifications of applying that licence is essential. ”

ASAPbio is not only focusing on preprints. They received a one million USD grant from the **Helmsley Charitable Trust** (helmsleytrust.org) in 2017 to form a **PubMed Central-style** (ncbi.nlm.nih.gov/pmc) central archive for preprints. However, this project was cancelled after some major changes in the preprint landscape meant that many of the goals of the project would be met elsewhere. Instead, the Helmsley Trust has allowed ASAPbio to use this grant for advancing transparency in peer review. Polka: “Earlier this year [in 2018] we published an **open letter** (asapbio.org/letter), now signed by hundreds of journals, that signals their commitment to publish the contents of peer review. All signees agree that publishing peer review reports (the contents of peer review, whether anonymised or not), would benefit the research community by increasing the transparency of the assessment process.”

ASAPbio is also working on a collaborative project, **TRANPOSE** (transpose-publishing.github.io), which aims to track the development of journal policies around publishing peer review.

Polka: “I don’t think that publishing the content of peer reviews will overhaul entire research workflows: Bringing them into the open, however, is a significant departure from established research practice. But on a practical level, I think recognition of preprints as a fully-fledged scientific communication channel will have the bigger impact.”

ASAPbio has a group of around 100 researchers as ambassadors, who not only share their ideas on preprints and open peer review, but also provide input and feedback to the organisation from their respective peer groups. It is a very bottom-up, community-oriented approach but that doesn’t mean Polka doesn’t see any room for commercial activities, in terms of providing services related to preprint publishing and open peer review: “I get worried when knowledge and information are treated as commodities, as proprietary items. If commercial entities are providing this kind of service, the challenge might be when the data is

not released publicly – this is inhibiting our ability to assess research.”

“ I don’t think the legal status of an entity necessarily reflects their commitment to an open infrastructure. Some of the most powerful opponents of open access have been non-profits. ”

“This happens when publishers are locking away abstracts or citations. So, on a fundamental level, these infrastructures should be publicly or community owned. But I don’t think the legal status of an entity necessarily reflects their commitment to an open infrastructure. Some of the most powerful opponents of open access have been non-profits. I think the entirely scholarly communication ecosystem is not functioning as a marketplace. The desire to publish in a prestigious place connected to career advancement prevents people from choosing the one that’s most suitable, and thus the most efficient way for them to publish their work.”

References and relevant links

- ▶ ASAPbio website: asapbio.org
- ▶ Resources about preprint licensing, created together with Creative Commons and PLOS: asapbio.org/new-licensing-resources
- ▶ Accelerating scientific publication in biology by Ronald D. Vale, Proceedings of the National Academy of Sciences Nov 2015, 112 (44) 13439-13446; doi.org/10.1073/pnas.1511912112
- ▶ Open letter about publishing reviews: asapbio.org/letter
- ▶ PubMed Central: ncbi.nlm.nih.gov/pmc
- ▶ TRANSPOSE project: transpose-publishing.github.io
- ▶ ASAPbio start meeting: asapbio.org/meeting-information
- ▶ The National Institutes for Health: nih.gov
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ PLOS : journals.plos.org
- ▶ Crossref: crossref.org
- ▶ UCSF: ucsf.edu
- ▶ BioRxiv: biorxiv.org

About Jessica Polka



Jessica Polka is the executive director of ASAPbio. Prior to this position, she was a postdoctoral research fellow in the department of systems biology at Harvard Medical School, mentored by Pamela Silver and co-mentored by Timothy Mitchison.

Polka received her BSc in Biology from the University of North Carolina at Chapel Hill (UNC-CH) and her PhD in Biochemistry from UCSF.

© Knowledge Exchange 2019

Title: Insights into the Economy of Open
Scholarship: A look into ASAPbio with
Jessica Polka, Executive director

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: September 13, 2018

DOI: [10.5281/zenodo.2586887](https://doi.org/10.5281/zenodo.2586887)

This report is available under a CC BY 4.0
licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info

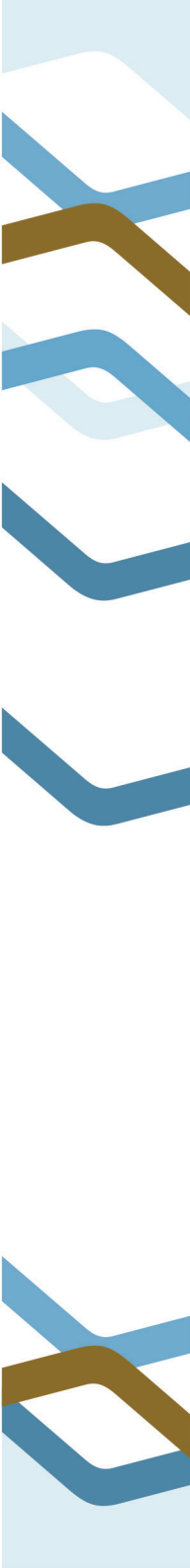
Insights into the Economy of Open Scholarship: A look into ScienceOpen with Stephanie Dawson, CEO



About ScienceOpen

ScienceOpen is an interactive discovery environment for scholarly research across all disciplines. It is freely accessible for all and offers hosting and promotional services within the platform for publishers and institutes. ScienceOpen is a privately funded startup company, owned by Alexander Grossmann and Tibor Tscheke. The main office is in Berlin, Germany with technical offices in Boston, USA and Budapest, Hungary. While the service is free for end users, publishers, institutes, and scholarly societies are charged either for full content hosting or for promotional services, which consist of the creation of 'collections', that is, curated landing pages collecting articles on a certain topic.

scienceopen.com



ScienceOpen: Business model

Key activities

- ▶ Discovery platform for open access articles
- ▶ Option to form cross-journal topical collections with advanced indexing and altmetrics options
- ▶ Conference poster publication platform



Organisation type

- ▶ Commercial startup
- ▶ Privately funded, owned by Alexander Grossmann and Tibor Tscheke



Key partners

- ▶ Publishers, scholarly societies, institutions
- ▶ UCL Press
- ▶ Member of CrossRef, ORCID, the Open Access Scholarly Publishers Association (OASPA), The International Association of Scientific, Technical and Medical Publishers (STM), and the Directory of Open Access Journals (DOAJ)



Revenue streams

- ▶ Private funding
- ▶ Fee paid by publishers, institutions, and scholarly societies
- ▶ Contract with UCL Press for mega journal



IP/Copyright

- ▶ All articles aggregated are open access (all licences allowed) – preprints/conference posters: CC BY 4.0
- ▶ Promotes open metadata and abstracts
- ▶ Own outputs: CC BY 4.0



Customers/users

- ▶ Publishers
- ▶ Scholarly societies
- ▶ Research institutions
- ▶ Individual researchers



Interview with Stephanie Dawson

The original idea behind ScienceOpen was to speed up science by removing the constraints of the traditional journal, and to experiment with digital tools to achieve this.

“ScienceOpen started in 2013, the year of the ‘mega journal’. **PLOS ONE** (journals.plos.org/plosone) appeared to be successful and changed the perception of what publishers can and should do in order to speed up the process of scientific communication. Post-publication collections suddenly became possible, for example, around Zika,” says **Stephanie Dawson**, CEO of ScienceOpen.

Many ScienceOpen employees had previous experience at big traditional publishing houses and experienced the digitisation process from the start, when it entailed nothing more than the digitisation of articles into pdfs. “The baggage of these legacy publishers was huge,” says Dawson, “and this was very frustrating. We couldn’t make colour images available online for technical reasons, there were always extra costs charged to the author, and the print version always remained the reference. We all felt that the full potential of going digital was not being used at all by the big legacy publishers.”

“ We all felt that the full potential of going digital was not being used at all by the big legacy publishers. ”

ScienceOpen set out to rethink this old-fashioned process. The idea was to get the research out as soon as possible with the community crowdsourced peer review happening afterwards in a mega journal platform environment. This mega journal would then be embedded in a larger discovery platform that would aggregate other information from **ArXiv** (arxiv.org), **PubMed** (ncbi.nlm.nih.gov/pubmed) etc so that every single article would be embedded in a much wider scientific context. “We hoped that our network would be able to communicate in a faster, digital and very transparent way about their research results because we also wanted to open up the peer review system.”

Although many researchers were very enthusiastic about the idea, ScienceOpen soon reached the level of innovation that authors were open to. “When it came down to publishing directly into our mega journal, they were suddenly worried about impact and reputation,” says Dawson. “In the end, it turned out that we had underestimated the implicit contract between publisher and researchers, which was still largely about reputation. This is a currency that you don’t get to offer as a brand new product.”

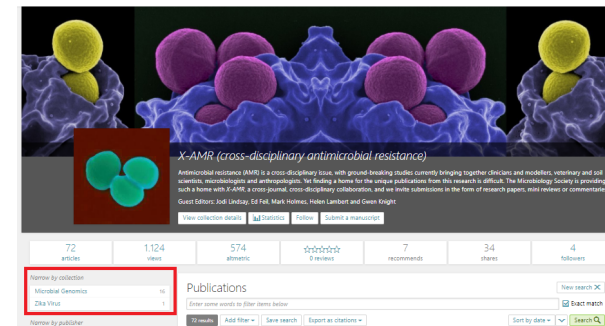
ScienceOpen decided to set its publishing ambitions aside, together with the idea of sustaining the platform by charging article processing fees. The decision was made to focus instead on offering services to publishers and societies that wanted to enhance the visibility of their niche journals by putting them in a broader context, including a search engine. The discovery platform remained in place, but an interactive overlay for user interaction and peer review was added. Dawson: “We focused on improving our overlay framework and services for researchers to convince publishers that our system could improve their visibility. The curation/collection

aspect is our unique selling proposition: we’ve made it possible to visualise different ways to curate content in a way that is not threatening to publishers or journal titles.”

For now, the main activities of ScienceOpen focus on the hosting of external content in a curated collection structure. Dawson: “We take a topical selection of articles. When working with publishers we can make a collection based on a single journal or, if we’re working with other entities such as scholarly societies, we can spread the net wider by making a selection from multiple journals.”

“For example, the **Microbiology Society** (microbiologysociety.org) created a collection on antimicrobial resistance that pulled from all of their journals, thus, creating a sort of ‘virtual’ journal. We promote these collections on our platform using banners and integrate it in our search function. The content remains hosted elsewhere though; we simply promote it via the metadata, linking back to the version of record on the publishers’ servers. It’s important to note that we only host and link to open

access articles, we are not interested in creating tools that will stop or even hinder the free flow of information.”



Example of a collection curated by the Microbiology Society

“Our plan to implement post-publication peer review has been pushed back for a while. I really wanted this to be an integral part of the contribution of ScienceOpen to ‘open’ but we haven’t yet been able to get researchers excited about this. We were the first ones to set up the **Crossref** (crossref.org) XML schema for post-publication peer review but it has not taken off yet, although it could be the next big opportunity.”

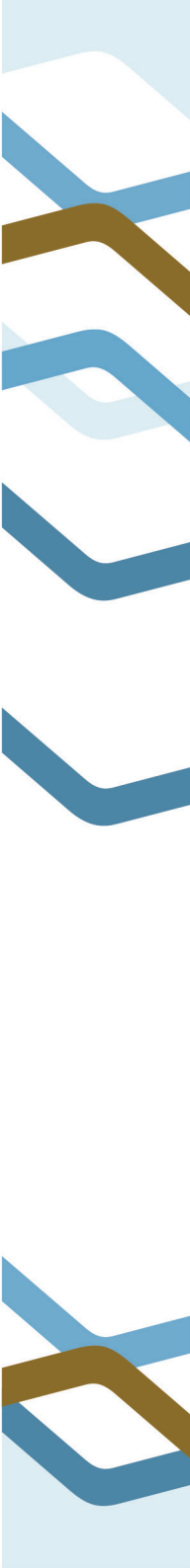
However, by engaging with **UCL Press** (ucl.ac.uk/ucl-press) in London, for which ScienceOpen is building a mega journal platform, ScienceOpen wants to revive this old ambition. “They really want authors to publish the preprint with the peer review happening in the open, and then publish the final peer reviewed versions. UCL is probably in a better position to make this work because they have enthusiastic researchers who will engage their peers and talk about open peer review. I think this grassroots approach can be more effective than if we, as a for-profit company, impose it on them.”

“ I think a grassroots approach towards post-publication peer review can be more effective than if we, as a for-profit company, impose it on researchers. ”

“It could be a great model for other institutions that want to start up a mega journal. This can be more impactful than just publishing in a repository or creating more small niche journals. Our set-up is perfect for overlay journals based on this peer review system. But we can’t do it on our own, so that’s why we consider our collaboration with UCL as a great demo project.”

Because the majority of its content is externally hosted, ScienceOpen relies heavily on the quality of the metadata it receives. The company is part of the **Metadata2020** (metadata2020.org) group, an organisation pushing for better quality metadata.

Dawson: “ScienceOpen offers an important indirect service to libraries, institutions, and individual researchers. We enable them to evaluate which publishers do a good job and which do a bad job in terms of discoverability of research, by providing good quality metadata, openness of references, and abstracts. These are all part of the metadata and should never be copyrighted, not only for the benefit of ScienceOpen, but for everybody’s benefit.”



“I feel it’s very much part of our mission to try to get the most out of the available digital tools,” says Dawson. “We really push all publishers we work with to care about the importance of metadata. Having your content tagged as open access, is something you need to do as a publisher, or you are doing a bad job at distributing the content. Authors are paying often high article processing charges (APCs) for this, and as a publisher you need to make this visible digitally. If you don’t do this, you deserve criticism! Publishers need to pay more attention to it. Algorithms such as that of **Unpaywall (unpaywall.org)**, which we use, are very good at detecting false positives but I am convinced that there are a lot of open access materials out there which are just not being picked up, simply because the metadata is inadequate. I fear this leads to a serious underestimation of the total amount of open access material that is actually available.”

“Inadequate metadata leads to a serious underestimation of the total amount of open access material that is actually available.”

Aside from its own hosting services, ScienceOpen is working with publishers to drive traffic to its version of record via their metadata. They get help to clean up and enrich their metadata and to add it to Crossref. ScienceOpen often works with publishers directly to convince them to improve their metadata.

Dawson: “Often it’s only a matter of changing or adding one metadata field! The smaller publishers who are struggling technically to add the right information to Crossref are more frustrating, either because they don’t have the technical knowhow or because they lack manpower. Then again, it’s not only the small companies. I know of some larger publishers who also just don’t care about getting the right metadata. For me this is all part of the same mission: to get better article and author information so that the entire research community can work more smoothly and more in concert with each other.

“We encounter the same reluctance when we discuss making abstracts openly available. We know that content with an abstract gets accessed and used up to ten times more than content without

one but not all publishers are actually taking the trouble to deposit abstracts with Crossref! If you don’t do this, which is the case for some major publishers, you are doing a disservice to your authors. Some are trying to monetise this essential part of research dissemination. I believe that making some research items open as default is a basic courtesy. Also, there are some cool artificial intelligence (AI) initiatives that can do such interesting stuff by mining the abstracts and references! I would like to see publishers take this more seriously.”

“If you don’t make abstracts openly available you are doing a disservice to your authors. Some publishers are trying to monetise this essential part of research dissemination.”

ScienceOpen offers open access hosting within its interactive discovery environment, including an individualised collection structure for a list price of 2,000 US dollars (USD) per year and a per article fee of 25 USD. Promotional collections can be booked for the same yearly fee with an article aggregation fee of five USD per article.

Dawson: “We always crosscheck with Crossref and the **Directory of Open Access Journals (DOAJ)** (doaj.org) for ‘new’ content, combining it with available metadata. If a publisher wants to go beyond the articles that we pick up and the metadata we can collect, they can also give us full metadata access to their catalogue. We can really work with them to improve visibility for individual articles, journals and so on. There are so many out there, it can be really hard if the publisher cannot differentiate its own brand from the competition.”

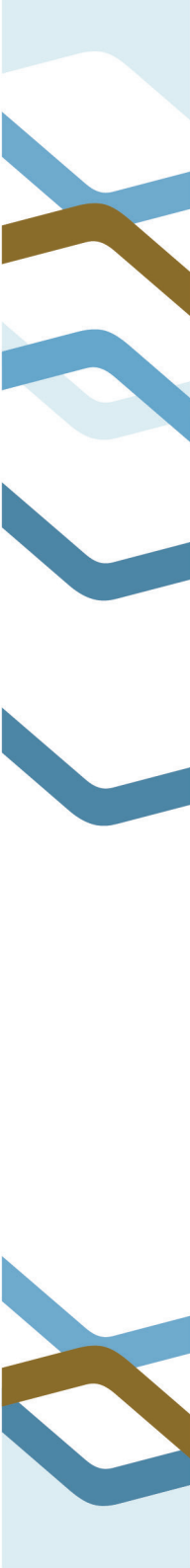
However, the service remains free for end users. In addition, researchers who want to create and edit their own topical collections can use the collection functionality. ScienceOpen even advises them about

curation and enhancement of these collections, for example, by suggesting additional articles or illustrations. Currently, 35,000 researchers use the service to promote their own work or to create a new collection for their field of interest. Dawson: “Publishers are also showing an interest in this, because for them it’s a very direct way to get feedback from researchers, both in terms of which collections are being created and which articles are being added to them.”

Above all, to become sustainable, ScienceOpen needs to ensure enough buy-in from paying customers. Right now, 150 paid-for collections are hosted on the platform, but in order for ScienceOpen to be both sustainable and to be able to invest more in development this number needs to increase. The company is trying to diversify revenue streams by licensing features to other projects, but thus far this has not been proven to be scalable.

For its end users ScienceOpen fosters a very open philosophy: it does not even require registration to use the discovery tools. Users only have to register if

they want to do something on the platform such as collection curation or peer review. From the beginning, ScienceOpen has been keen to promote the use of **ORCID** (orcid.org) (the persistent identifier for researchers). ScienceOpen asks people to edit and augment their ORCID profiles, instead of editing user profiles on the platform. Everything is imported from the ORCID servers.



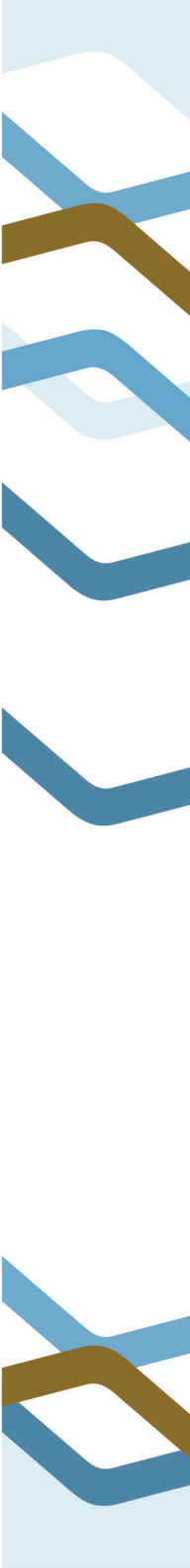
Dawson: “I think this way of working illustrates our philosophy nicely! ORCID is an excellent digital tool and it makes research dissemination more efficient and interoperable between systems because there is less confusion about who wrote what. The extra step might have deterred some people in the beginning, but we still think it’s worth it. Also, it’s more efficient for us as we don’t have to manage all this personal information.”

“ScienceOpen is a for-profit company. We do occasionally get criticised for that. However, I think for-profits have a place in the service provider infrastructure. Universities work with commercial companies in so many different areas. I don’t see it as a big issue. But it has made it difficult for us to work with libraries, for example. They work with commercial publishers all the time, but when they start thinking about open access infrastructure they project their bad experiences with those publishers onto us.

“There’s certainly wariness about potential data or vendor lock-in. But programming for yourself is usually very expensive, so I don’t think it’s possible to have all your services built in-house based on not-for-profit open source infrastructure.”

“We try to be very cautious about not creating any data lock-in ourselves, but we need to get even better at respecting the work that goes into hand-curated collections. If an editor wants to move their collection they can export the metadata, but only for 200 items at a time. We are working on a system so that researchers will always have the assurance that they can download the metadata of their collections in its entirety, hassle-free. We don’t have contracts with researchers but that’s one of the places where we have to come up with a good technical solution so that researchers know they’ll always be able to take their data with them.”

Although services similar to ScienceOpen, such as **Scopus** (scopus.com) and **Web of Science** (wok.mimas.ac.uk), cover a large part of the market, Dawson doesn’t think they represent direct competition because these are paywalled services. However, she believes **Google Scholar** (scholar.google.co.uk) is much more of a threat because of its powerful algorithm: “On the other hand it can be argued there are weaknesses in their model – because they are essentially a ‘black box’.



“I really see open scholarship as being committed to shifting the research paradigm to a more collaborative mode in the digital space, with the brand new digital tools that we now have available to us. We know how to solve problems faster with open science. For me, a great example is the H1N1 swine flu outbreak in 2009: the **American Centers for Disease Control and Prevention (CDC)** (cdc.gov) coordinated efforts, uploading the complete gene sequences to a public database to enable global cooperation. The goal was not to get credit but to get their results out there as fast as possible, to get a vaccine as soon as possible. For me, that is really what open science is about. Why can't we all work more collaboratively, globally, with the digital tools that we have? How do we need to change the reward system, change scholarly structures, so that scholars can solve the big issues?

“We have the tools, but we haven't figured out how to get everybody on board yet. It will require a cultural shift that will take some time – but with ScienceOpen, we are trying to make a contribution,” concludes Dawson.

References and relevant links

- ▶ ScienceOpen website: scienceopen.com
- ▶ PLOS ONE: journals.plos.org/plosone
- ▶ ArXiv: arxiv.org
- ▶ PubMed: ncbi.nlm.nih.gov/pubmed
- ▶ Microbiology Society: microbiologysociety.org
- ▶ Crossref: crossref.org
- ▶ UCL Press, London, UK: ucl.ac.uk/ucl-press
- ▶ Metadata 2020: metadata2020.org
- ▶ Unpaywall: unpaywall.org
- ▶ DOAJ: doaj.org
- ▶ ORCID: orcid.org
- ▶ Scopus: scopus.com
- ▶ Web of Science: wok.mimas.ac.uk
- ▶ Google Scholar: scholar.google.co.uk
- ▶ Centers for Disease Control and Prevention: cdc.gov

About Stephanie Dawson

CEO



Stephanie Dawson grew up in northern California and studied biology at Yale University. She then worked at the labs of Susan Parkhurst at the Fred Hutchinson Cancer Research Center in Seattle, WA and Ralph Rupp, at the MPG Friedrich Miescher Laboratory, Tübingen, Germany before changing fields and getting a PhD in German literature from the University of Washington under Jane Brown. From 2001-2012 she worked in various positions at the academic publisher De Gruyter in Berlin in the fields of biology and chemistry in both journals and book publishing. In 2013 she joined the ScienceOpen management team as CEO.



© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A look into ScienceOpen with Stephanie Dawson, CEO

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: November 15, 2018

DOI: [10.5281/zenodo.3084650](https://doi.org/10.5281/zenodo.3084650)

This report is available under a [CC BY 4.0 licence](https://creativecommons.org/licenses/by/4.0) creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info

Insights into the Economy of Open Scholarship: A look into HRČAK with Jadranka Stojanovski, University of Zadar/Ruđer Bošković Institute, HRČAK Advisory Board



DFG



Jisc

CNRS

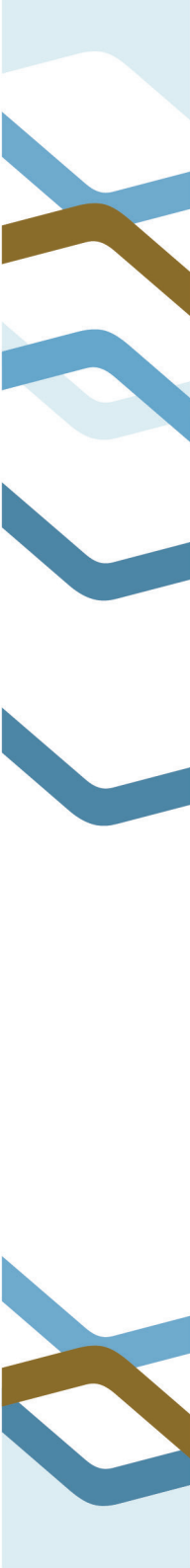
SURF

Ministry of Higher
Education and Science
Danish Agency for Science
and Higher Education

About HRČAK

HRČAK is the central portal of Croatian scientific, professional and popular open access (OA) journals. HRČAK offers open access to almost 500 journals, 300 of which are scientific peer reviewed publications, mostly in Croatian. Around 60% of the journals are in the humanities and social sciences (HSS) subjects. The portal, which was government-funded in the first three years of its production, is now an integral part of the University of Zagreb University Computing Centre (SRCE) and also receives occasional funding from international grant-funded projects. Most of the journals do not ask authors to pay article processing charges (APCs).

hrcak.srce.hr



HRČAK: Business model

Key activities

- ▶ Journal platform for 500 Croatian open access journals, of which 300 are peer reviewed scientific and professional journals



Organisation type

- ▶ Non-profit
- ▶ Staff: 1.2 FTE



Key partners

- ▶ Academic and research libraries
- ▶ University of Zagreb Computing Centre
- ▶ Government
- ▶ Universities
- ▶ Journal publishers



Revenue streams

- ▶ HRČAK journals: government funding, institutional funding, a few journals experiment with author fees
- ▶ University Computing Centre (SRCE): staff and infrastructure
- ▶ Government funding (first three years)
- ▶ Grant funding



IP/Copyright

- ▶ Code base is not open
- ▶ Articles: open access, some Creative Commons licensed



Customers/users

- ▶ Researchers/authors
- ▶ Journal editors
- ▶ Publishers
- ▶ Government
- ▶ Citizens



An interview with Jadranka Stojanovski

Scholarly publishing in Croatia has always had its idiosyncrasies: There are no significant commercial players active in the field and most journals are published and funded by universities, professional associations, and societies. These collect some funding via print subscriptions but, unlike the textbook publishing industry, where commercial publishers play a role, there have never been real market forces at play.

“When international journals started being published in digital and online versions in the early nineties it was obvious that Croatian publishers weren’t following this migration at the same pace. There was a discrepancy between the relatively large number of scholarly journals in Croatia and the small proportion available online. These journals had neither enough technical know-how nor the funding needed to hire professional staff who could provide production, sales, marketing, and administrative support,” says **Jadranka Stojanovski** from the HRČAK Advisory Board.

In 2005 a small group of librarians and information specialists wrote the initial project proposal to build a common platform for Croatian journals, HRČAK (hamster in English), which would enable journal editors to publish their content online free of charge.

As a result of a long partnership, the proposal for the HRČAK project was successfully submitted by the **SRCE** (srce.unizg.hr/en/university-zagreb-university-computing-centre-srce). “When we started with HRČAK, we had the goal of attracting at least 50 journals to use the portal. Nowadays we have almost 500! Besides around 300 peer reviewed journals, HRČAK also collects professional,

popular and student publications, as well as some trade and industry journals.

“HRČAK is a good example of a fruitful collaboration between journal editors, librarians and ICT specialists,” says Stojanovski.

“ Currently, HRČAK is still hosted, maintained, and developed by SRCE and it is a good example of a fruitful collaboration between journal editors, librarians, and ICT specialists. ”

Stojanovski: “We’ve always collaborated well with journal editors and our model, providing them with free infrastructure to publish their journals in open access at no cost, was widely accepted. HRČAK became popular quickly, measured by publicly available usage statistics. For some journals our platform offers the only digital version available but, for the majority, editors are providing to HRČAK the same content as published on their websites.”

Government support for journals has always been a tradition in Croatia, especially for HSS research, which is mainly published in Croatian. The entire scholarly community in Croatia is only around 11,000 researchers, and without some form of support these journals would have severe sustainability issues. After an evaluation process journals are assigned different government-issued subsidies ranging from 800 EUR to 32,000 EUR per year, which is about 80% of their costs.

“We are also looking for projects and grant funding schemes. Partnership and collaboration with bigger consortia such as **OpenAIRE** (openaire.eu) and **OPERAS** (operas.hypotheses.org) are crucial for us. Also, we would like to establish a collaboration with **SciELO** (scielo20.org) (Scientific Electronic Library Online) since we find many similarities between HRČAK and the SciELO platforms.

“It is important to stress that the platform was open access from the start even when there was no formal policy in place yet. When the advantages of open access became obvious we made it mandatory,” says Stojanovski.

“Very important for us was the support from the Ministry of Science and Education and their publishing committee; publishing on HRČAK (and thus in open access) became a criterion for journals to receive government funding.

“Together with the ministry’s publishing committee, which is in charge for journal subsidies, we are continuously improving the criteria for evaluating journals.

““ What is important to stress is that the platform was open access at its earliest stage – even when there was no formal policy in place yet. ””

As a stated goal is to emphasise the value of locally relevant research, HRČAK is also working with publishers and editors to raise the quality of Croatian journals, ensuring that editorial policies are up to standard and encouraging them to follow new trends in scholarly publishing.

Stojanovski: “What is quite exceptional is that the majority of journal editors are very much in favour of open access. They realise it’s the most efficient way to attract new audiences and to guarantee their authors the biggest possible visibility and impact. We’re encouraging them not only to open up the article, enabling wide usage of **Creative Commons** (creativecommons.org) licences, but also to open up the underlying research data. Also, we’re currently promoting new approaches such as open peer review as well, although we’re still waiting for the first journal to employ it.

“HRČAK still reflects our 2005 vision, which was mainly about providing a common infrastructure for journals to publish their content online at no cost for them or for the reader. But this doesn't respond to current journals' needs anymore. The new vision should be to offer more in-depth services such as modern editorial standards and technical innovations. In the future, HRČAK could be a publishing platform that encompasses all stages of scientific publishing from submission and the peer review process to publication. We would like to support the editorial process in a more integrated way. This could be achieved, for example, by the full integration of Open Journal Systems (OJS, openjournalsystems.com) or by an appropriate development of other editorial modules. We're already offering OJS to editors and it is quite easy to exchange the metadata with HRČAK, but I think it would be better to have both systems completely integrated.

“Applications or plug-ins helping authors and editors to create XML versions of articles for free would be welcomed by our journals. I believe this approach will ensure best editorial practices and research integrity, encourage new types of peer

review, and so on. Another issue that we face now is that many journals publish the same content on their websites as they do through our platform – which doubles the effort and makes it difficult to aggregate usage statistics. We should be evolving towards a full-blown publishing platform.

“Our ultimate goal is to enable all Croatian journals to become high quality open access publications, with all research data available and open peer review in place, improving scholarly communication by use of the available technologies. We're also looking into enhanced publications with dynamic, multilayer, interactive, multimedia content. On the technical side, we're focusing on machine readability of the articles, linked data, open research data, variety of formats (beyond PDF), global author identification (**ORCID**, orcid.org), persistent identifiers such as digital object identifiers (DOIs) and so on. When everything is open access and machine readable, publishers will be able to provide further added value, more functionalities and advanced services.”

“ Our ultimate goal is to support all Croatian journals to become high quality open access journals, with all research data available, with open peer review in place, improving scholarly communication by use of the available technologies. ”

In recognition of changing research workflows, and to stimulate them, HRČAK also wants to support other publication types besides journals, such as conference proceedings, books and educational materials. “We don't have enough resources to develop HRČAK-like platforms for conference proceedings and other types of publications as fast as we'd like to,” says Stojanovski, “and we're also lacking in training resources – we haven't been able to train editors and authors as much as we'd like to. In addition, we need to pay more attention to emerging topics such as research integrity and other ethical issues. I also believe that we should evolve towards open peer review, with optional

disclosure of a reviewer's identity. Disclosure could be tricky in such a small research community, but openly available reviews could improve the peer review process, raising quality standards.

“At this point, we don't have sustainable funding available for HRČAK, and SRCE uses its regular staff for HRČAK tasks, while working on other projects and services. The members of HRČAK advisory board are all volunteers, and their engagement depends on how much time they can set aside.” HRČAK used grant funding from the EU project OpenAIRE to implement more advanced features, but Stojanovski thinks that they would benefit from having a dedicated team working exclusively on development. There are different publishing platforms in Europe working separately, so maybe more international collaboration and cohesion could be a way to proceed faster.

Stojanovski: “We also see competition developing: Certain publishers are approaching Croatian journals and taking over the publishing role, but we're unsure about outcomes. On the one hand, the publishers might demand the introduction of APCs, which is

more of a source of income for publishers than of support for editors in Croatia, who work hard to provide reliable content. On the other hand, the reputation of the publisher makes the odds of inclusion in popular databases like Scopus and Web of Science much more likely, compared to an unknown Croatian publisher, even if content of the journal is identical. Our biggest strength is that our staff are very knowledgeable and competent. We have an excellent relationship with our journal editors and, because the community is relatively small, we can reach most members relatively easily. This direct contact is of great value to us.”

Although the predominant scholarly publishing culture in Croatia is still not based on authors' fees, there are now some HRČAK journals introducing APCs. Stojanovski: “Until recently, Croatian editors and publishers were unanimously against APCs because they essentially transform an open scholarly publishing system into a business, which has nothing to do with science itself.”

“Until recently, Croatian editors and publishers were unanimously against APCs because they considered them crucial in transforming an open scholarly publishing system into a business, which has nothing to do with science itself.”

“Firstly, it is economically unfeasible for authors or funders from countries with low levels of research and development (R&D) investment to pay them, but it's also not in line with their journal publishing philosophy, which was never commercially oriented.

“Recently, some HRČAK journals introduced APCs not only because it's a way to get more (and more sustainable) funding, but also because they want to align themselves with the mostly author fee-based OA scholarly publishing in Western Europe. There are many small non-profit publishers in the world, some of them offering a more advanced approach towards publishing, but I am still wary of large

corporations controlling academic publishing and their influence. I believe that ‘Open Scholarship’ will be harmed if commercial publishers dictate its course. This development should be defined by the needs of the research community and society in general, and governed by openness. I find that discussions on open access focused exclusively on big publishers’ business models, equating open access with APCs, are unproductive. We should be discussing the evolution of scholarly publishing into creating more efficient tools for sharing ideas, methods, and research results and shift away from paper-centric publications. The prevalent pro-APC discourse in open access publishing is a demonstration of how commercial interests are influencing the scientific publication process.

“ I believe that ‘Open Scholarship’ will suffer definite harm if commercial publishers dictate the direction it will go. This direction should be defined by the needs of the research community and society in general, not by commercial interests. ”

“Publishers can play a role in developing state-of-the-art innovative services, which they can sell as a product to enhance scholarly communication. In that way, public/private partnerships can offer added value and make sense economically. However, they should not be allowed to lock in research articles or data. Publishing research results in open access incurs some costs which need to be covered, but in a system where the content is provided by authors/researchers and is evaluated by peer reviewers/researchers free of charge, the money should be invested in a variety of forms, formats, and media for the appropriate presentation of research results, accompanied by diverse services and with good text and data mining (TDM) tools available.

“When author fee-based publishing became the dominant model, I really had my doubts and I felt like a Grinch stole open access!” Stojanovski remains an outspoken advocate against author fee-based open access publishing: “Open access is, for sure, the future of scholarly publishing, but we need to support a variety of approaches and not-for-profit business models equally.

“With HRČAK, we have found that investing a relatively small amount of funding, supported by enthusiasm and lot of volunteer work, leads to a bigger common good. Open scholarship is about bringing back scholarship to the researchers, it’s about improving scholarly communications in general, it’s about bringing value to society – not only to advance the researcher’s career. It’s also about using new and digital tools and channels to disseminate research beyond the formats inherited from the printed world.

“ With HRČAK, we have found that investing a relatively small amount of funding, supported by enthusiasm and lot of volunteer work, leads to a bigger common good. ”

References and relevant links

- ▶ University of Zagreb Computing Centre (SCRE): srce.unizg.hr/en/university-zagreb-university-computing-centre-srce
- ▶ SciELO (Scientific Electronic Library Online): scielo20.org
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ Open Journal Systems (OJS): openjournalsystems.com
- ▶ ORCID: orcid.org
- ▶ OpenAIRE: openaire.eu
- ▶ OPERAS: operas.hypotheses.org

About Jadranka Stojanovski



Jadranka Stojanovski is an assistant professor at the University of Zadar, Department of Information Sciences and research librarian at Ruđer Bošković Institute in Zagreb, a member of the Commission expert group on National Points of Reference on Scientific Information, the OpenAIRE NOAD for Croatia, and an Open Science advocate. She was involved in the creation of the OA information infrastructure, including the Croatian Scientific Bibliography CROSBI, Who's Who in Science in Croatia, ŠESTAR repository of Scientific Equipment, HRČAK and DABAR, enabling open access to the knowledge created by the Croatian academic and research community.

Stojanovski's research is in the field of scholarly communication and Open Science, research integrity, and next-generation metrics to assess research output.

© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A look into HRČAK with Jadranka Stojanovski, University of Zadar/ Ruđer Bošković Institute, HRČAK Advisory Board

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: October 16, 2018

DOI: [10.5281/zenodo.2586868](https://doi.org/10.5281/zenodo.2586868)

This report is available under a [CC BY 4.0](https://creativecommons.org/licenses/by/4.0) licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info

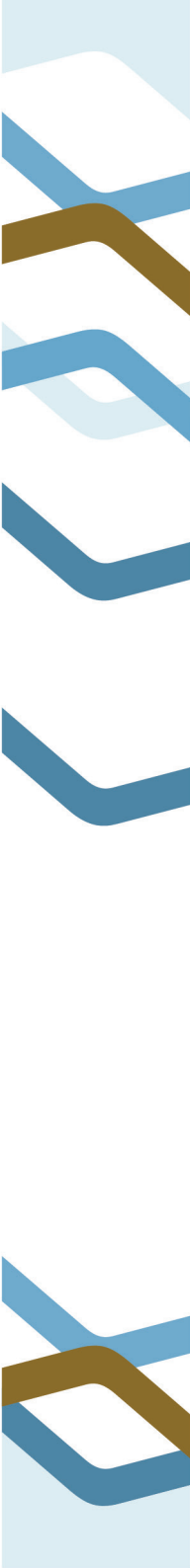
Insights into the Economy of Open Scholarship:
A look into Helsinki University Press with
Leena Kaakinen, publishing director



About Helsinki University Press (HUP)

Helsinki University Press is a startup open access (OA) university press. It is owned by the University of Helsinki (UH), and the management of the publisher's operations is shared between the **Helsinki University Library** (helsinki.fi/kirjasto/en/home) and **Gaudeamus** (gaudeamus.fi/in-english), a publishing house for non-fiction literature in Finnish. HUP plans to publish its first books in 2019. In its startup phase, HUP is funded by the University of Helsinki and it is not charging any processing fees. HUP is currently conducting talks with other universities and funders to investigate a consortium model that will allow it to cover the publishing costs centrally, rather than charging processing fees to individual authors. Currently, HUP has three staff members and an academic advisory board.

hup.fi



HUP: Business model

Key activities

- ▶ Open access press
- ▶ No author fees
- ▶ Focus on books (monographs and edited volumes)
- ▶ Investigating journal publishing



Organisation type

- ▶ University press startup
- ▶ Funded by UH, prospecting new sources of income
- ▶ Staff: Two full-time equivalent (FTE)



Key partners

- ▶ University of Helsinki
- ▶ Helsinki University Library
- ▶ Gaudeamus Publishing
- ▶ Ubiquity Press
- ▶ Universities in Finland (ongoing)
- ▶ Other university presses, mainly in UK and Sweden



Revenue streams

- ▶ Funded by University of Helsinki
- ▶ (Potentially) article processing charges (APCs) for non UH authors
- ▶ (Potentially) sales of print books
- ▶ (Potentially) library consortium model



IP/Copyright

- ▶ Outputs: open access default: CC BY 4.0
- ▶ More restrictive open licences possible



Customers/users

- ▶ UH researchers
- ▶ Researchers at other institutions in Finland



Interview with Leena Kaakinen

UH has long dreamt of a press that would publish in English for an international audience and it wanted to promote open access publishing. With Helsinki University Press (HUP) it is aiming to combine both, using the publishing expertise of Gaudeamus, the university's press for Finnish publications, and the experience of Helsinki University Library.

“We believe the press will benefit from these different types of expertise,” says **Leena Kaakinen**, publishing director at the new press. “Our publishing staff is very experienced in academic publishing. We have strong experience in author-facing services such as copy-editing and organising peer review, while at the same time we have the benefit of working with the library and being able to use their scientific knowledge as well.”

“We've also consulted with external experts such as **UCL Press** (ucl.ac.uk/ucl-press) in London (UK), **Stockholm University Press** (stockholm.universitypress.se) in Sweden, and **Ubiquity Press** (ubiquitypress.com) also based in London. We believe that, by setting up this press, we are answering the need for more good quality open access publishing channels. In addition, because we are well connected internationally we are in a good position to establish ourselves. We also have an excellent academic advisory board. They are very motivated and a very good resource for the press.”

A crucial process during the establishment of the press has been intensive consulting sessions with researchers from all disciplines. Before designing policies and workflows, HUP discussed researchers' expectations, hopes, and potential issues. Publishing cultures, needs and attitudes towards open access vary considerably between different fields of research and also between researchers at different stages of their careers.

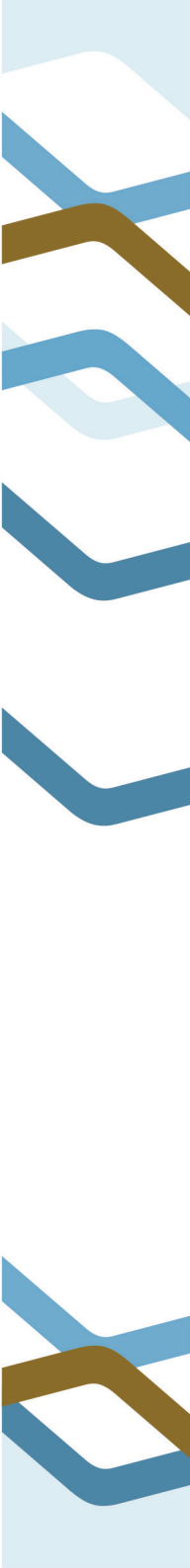
“ For some researchers it is hard to see the benefits of open access publishing in their everyday work, while for others this is clear. ”

“In some fields the discussion is already happening, mainly via open access channels, whereas in others the traditional model of publishing is still very strong. For some researchers it is hard to see the benefits of open access publishing in their everyday work, while for others this is clear, and they have concrete examples of how open access has been beneficial for them,” says Kaakinen. “One very important finding of these consultation sessions is that many researchers find the current reward system confusing. On the one hand they are encouraged or even compelled to publish open access, but on the other hand the reward system forces them to publish in high impact, often non-open access publishing channels if they don’t want to ruin their career... or at least, that is their perception.

“To attract authors, proving HUP’s potential impact will be a challenge. The current funding cycle still encourages researchers to publish in high impact journals, which are usually long-standing closed access or hybrid journals. This slows down open access in general but has a particularly big influence on us as a startup press. We don’t have a proven impact yet.”

“Initially, we will therefore aim for researchers who are already in favour of open access. Finding these early movers and providing good services for them is really crucial in this startup phase. I hope our experience with Gaudeamus will help here. To convince the others, we’ll have to focus on dissemination and visibility – we really need to provide added value there.”

“We provide digital dissemination through all open access channels relevant to the field including indexing and active marketing to improve visibility. We use social media and other channels, and with each book we seek the relevant communication channels for that particular field and communicate through them.”



For its infrastructure, that is the platform on which the press will run, HUP is relying on its membership of the **Ubiquity Partner Network** (ubiquitypress.com/site/partners), run by the London-based open access publisher Ubiquity Press. “We are such a small actor; for us to build our own platform would be by far the most expensive option. That’s why we are members of the Ubiquity Partner Network,” says Kaakinen. “It is useful for us to be part of a larger network and to buy the ready made platform services from them. It is a commercial company but they are very transparent about it and the contractual guarantees they give in order to avoid data or vendor lock-in are sufficient for us. Of course, that was crucial – there are other options around for this so we want to make sure that we have enough flexibility.”

“ We are such a small actor...it is useful for us to be part of a larger network and to buy ready made platform services from [Ubiquity Press]. ”

For the moment, HUP is mainly text-based with the option of augmenting publication with audio-visual aspects. Discussions are ongoing with Helsinki University Library to collaborate on establishing a shared data platform that will allow storage, archiving and dissemination of research data underlying HUP publications.

The central licensing policy for HUP will be **Creative Commons** (creativecommons.org) Attribution 4.0 (CC BY 4.0), but the publisher will allow some flexibility if requested by the author: “We will allow more restrictive licences such as CC BY-NC 4.0 (Creative Commons Attribution Non-Commercial) and CC BY-ND (Creative Commons Attribution No Derivatives) if the author really wants it. Researcher attitudes towards licensing vary. Many researchers don’t know very much about open access publishing yet, let alone open licensing. We really have to train them about what open licences mean. Some researchers worry that a third party could make commercial use of their work and want to use a non-commercial licence, for example. Also, some have expressed other worries about CC BY,

because they fear their work will be taken out of context and cited in a misleading way, and therefore they would prefer to use the ‘no derivatives’ clause.”

“ We really have to train researchers about what open licences mean. ”

As with every startup, designing a sustainable financial model has been a challenge. In its first phase, HUP will be supported entirely by the university. Publication will remain free of charge for UH researchers and, for the moment, publishing fees for external authors will be covered by the university funding as well. HUP is currently investigating funding models to achieve sustainability without having to ask for publishing fees. One of the options is to establish a consortium model with other institutions, to attract the main funders in Finland.

“ We don’t want publishing fees to be a burden for authors, which will be the case if there are no clear paths for them to apply for institutional funding. ”

“We feel that charging fees to authors is a problematic way to cover the costs for open access publishing. We don’t want publishing fees to be a burden for authors, which will be the case if there are no clear paths for them to apply for institutional funding. This became very clear during our consultation rounds with the researchers. If there are any publishing fees they should be paid by the institution or the funder,” says Kaakinen.

“ Authors are always puzzled as to how and where they should apply for fees. But, as it concerns all open access publishing venues in Finland, we need to collaborate in order to find a solution. ”

“Maybe this can even be arranged via us, the publisher, where we take care of the application process. I think this is a universal issue – authors are always puzzled as to how and where they should apply for fees. But, as it concerns all open access publishing venues in Finland, we need to collaborate to find a solution.”

HUP would like to obtain a proportion of its income from the sales of its publications as e-books or in print form, but as all outputs will be free to download it’s not clear what the revenue from that will be.

Kaakinen: “Depending on how well we manage to secure revenue from other funders and institutions we’ll investigate how to cover the fees for non-HU publishers. We’re having these negotiations right now and it’s not easy to project what will happen.”

As monograph publishing requires one-off funding instead of the continuous support that an open access journal demands, HUP will focus on books in its startup phase.

Looking further ahead, the press is currently investigating how to add journals to its portfolio: “I hope that in 2019 we can start with between four and eight books as a way to officially launch the press. Eventually, we’re aiming to having 20-40 books per year in our portfolio. Although in this startup phase we’re focusing on books, we will have a portfolio of journals. These are society journals, and working with them will be a way to partially cover our costs,” says Kaakinen. “But the talks are still going on about that and we’ll have to investigate whether we can take more journals on board.”

“The main issue is that they need long-term funding, and the existing funding landscape will probably change here in Finland because of the shift towards open access.” Finland’s biggest funder, Academy of Finland, is supporting the open access publishing initiative **Plan S** (coalition-s.org), and there are questions as to whether this will influence HUP’s funding model.

Kaakinen: “We really need to communicate why open access is so important, and we need to tailor it to researchers. One size does not fit all because there are different needs in different fields of research. Funders and policymakers have a big role to play; the current funding and evaluation systems are confusing and do not necessarily steer researchers towards more open scholarship. Open access mandates and policies need to be accompanied by concrete measures in order for them to be effective. There will always be costs incurred with open access publishing, so this will need to be addressed. There is no clear action plan yet [in December 2018, after the date of this interview, Academy of Finland actually started a **consultation session (aka.fi/en/about-us/media/press-releases/2018/open-consultation-to-gather-input-on-plan-s-implementation/)** about Plan S implementation in Finland, Gwen Franck] – so there’s a lot of uncertainty about how the funding streams will be redirected and what will happen during the transition. It’s difficult to predict what the consequences of these policy changes will be, for us as a startup open access publisher.”

“ Open access mandates and policies need to be accompanied by concrete measures in order for them to be effective. ”

References and relevant links

- ▶ HUP website: hup.fi
- ▶ Gaudeamus: gaudeamus.fi/in-english
- ▶ Helsinki University Library: helsinki.fi/kirjasto/en/home
- ▶ UCL Press: ucl.ac.uk/ucl-press
- ▶ Stockholm University Press: stockholmuniversitypress.se
- ▶ Ubiquity Press: ubiquitypress.com
- ▶ Ubiquity Partner Network: ubiquitypress.com/site/partners
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ Plan S: coalition-s.org
- ▶ Announcement of Plan S consultation session by Academy of Finland: aka.fi/en/about-us/media/press-releases/2018/open-consultation-to-gather-input-on-plan-s-implementation

About Leena Kaakinen

Publishing director



Leena Kaakinen is the publishing director at Gaudeamus and is also the publishing director for Helsinki University Press.



© Knowledge Exchange 2019

Title: Insights into the Economy of Open
Scholarship: A look into Helsinki University
Press with Leena Kaakinen, publishing director

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: November 22, 2018

DOI: [10.5281/zenodo.2843316](https://doi.org/10.5281/zenodo.2843316)

This report is available under a CC BY 4.0
licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info

ImpactStory.

Insights into the Economy of Open Scholarship:
A look into Impactstory with
Heather Piwowar, co-founder



About Impactstory

Impactstory is a non-profit dedicated to making scholarly research more open, accessible, and reusable. It offers free services including Unpaywall, Impactstory Profiles, and Depsy. The company is based in Vancouver, Canada. Its mission is to change the focus of the scholarly reward system to value and encourage web-native scholarship. The organisation receives or has received grant funding from The Open Society Foundation, The Alfred P. Sloan Foundation, The National Science Foundation (NSF), Clarivate Analytics, and Arcadia. Aside from this, Impactstory earns income from selling the Unpaywall Data Feed to 12 high-profile customers and from a smaller subcontracting grant with the University of Texas at Austin.

impactstory.org



Impactstory: Business model

Key activities

- ▶ Unpaywall: free version and data feed (paying service)
- ▶ Depsy
- ▶ Impactstory Profiles
- ▶ Research/consultancy



Organisation type

- ▶ Non-profit
- ▶ Two full-time equivalent (FTE) staff (+ One additional hire in December 2018)



Key partners

- ▶ Funding agencies
- ▶ University of Texas at Austin
- ▶ Data providers



Revenue streams

- ▶ Earned income from Unpaywall data feed
- ▶ Grant funding
- ▶ Subcontracting/consulting for University of Texas at Austin



IP/Copyright

- ▶ Code base: open source
- ▶ Data: depends on status of the data provided



Customers/users

- ▶ Publishers, funders, institutions, companies
- ▶ Individual researchers



Interview with Heather Piwowar

Since its foundation in 2011, Impactstory has released a range of services to help researchers measure the impacts of their research outputs. What started out with an attempt to build an author profile system intended to encourage altmetrics and offer incentives for sharing research, has now become a well-established organisation with Unpaywall as its most well-known product.

“Essentially, Impactstory is a two-person non-profit,” says co-founder **Heather Piwowar**. “My co-founder Jason Priem and I met at a hackathon during our PhD years. Our initial idea was to build a profile system for researchers that offered metrics and incentives for sharing. A small grant from **Sloan (sloan.org)** and a collaboration with **Duke University (duke.edu)** allowed us to work full time on this project. However, despite people loving our profile project on social media, we only got about 20,000 people to actively use it, which was not enough for it to be sustainable. Even after we experimented with asking a fee for it, it became clear that we would not be able to continue to run the service.”

Although the research profile service did not take off, Impactstory procured some other grants to fund **Depsy (depsy.org)**, a tool for researchers to showcase open source software they have created and to add value to it, promoting credit for software as a fundamental building block of science.

The organisation is also acting as a subcontractor for the **University of Texas at Austin on a project to improve research software (ischool.utexas.edu/tags/impactstory)** by helping its creators get proper credit for their work. Another project Impactstory worked on was ‘Open Heroes’, a badge system to reward researchers who made their work openly available.



Piwowar: “While developing these services, we realised there was no good all-encompassing way to determine whether papers were actually open access. A lot of services we relied on (such as **CORE (core.ac.uk)**, **Open Access Button (openaccessbutton.org)**, and **OpenAIRE (openaire.eu)**) did not, at that time, deliver the kind of results we required, at the very high volume we needed – namely a clear ‘yes’ or ‘no’ that was actually accurate. That’s why we decided to build our own code based on the **BASE (base-search.net)**, **PubMed (ncbi.nlm.nih.gov/pubmed)**, and **ArXiv (arxiv.org)** APIs [currently, Impactstory is no longer using the BASE API, Gwen Franck]. For us, precision is key: if we say something is open access, it definitely is. We don’t want surprise embargoes where a paper that is labelled as ‘open’ turns out to be embargoed after all, as happens with many repositories. With us, the link will always lead to the open access version of the paper.

“ For us, precision is key: if we say something is open access, it definitely is. ”

“But as it turned out, the **Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH (openarchives.org/pmh)** standard does not have a single way to ensure that the version directed to is open. That’s why many repositories do not show this data accurately and that’s why some aggregators don’t display the correct information.”

However, even after building their own code, Impactstory still had issues, either because the data was not accurate enough, it was not delivered fast enough, or because they could not call the API often enough. In one case, it was difficult to get permission for commercial reuse and this was problematic, because they intended to sell the service. Starting from this initial product, Impactstory decided to spin it off to create its own API.

The Unpaywall API finds open access content in many places, including using data from open indexes like **Crossref (crossref.org)** and the **Directory of Open Access Journals (DOAJ (doaj.org))**, but the majority of the open access content comes from independently monitoring over 50,000 unique online content hosting locations. This new service, the ‘Unpaywall Data Feed’, quickly took off as libraries started to integrate it. They built a browser extension and started doing regular data snapshots. This is now Impactstory’s best-known service.

Data Feed: Changefiles

The Data Feed changefiles show all the changes in the Unpaywall database over time. They are provided for subscribers to the [Unpaywall Data Feed](#). A new file is added every Thursday. Files use the same [schema](#) as the [REST API](#) and [database snapshot](#).

This list is also available via a [JSON endpoint](#) for programmatic access.

API key required

Paste your API key here

ACCESS CHANGEFILES

Unpaywall Data Feed

Apart from the API, which gets about three million calls a day, Impactstory makes a data dump available for commercial and non-commercial use every six months. It also has a weekly dump showing access status changes of individual papers, based on their **Digital Object Identifier (DOI) (doi.org)**. This weekly dump is a paying service, for which Impactstory currently has 16 high-profile paying customers, including Elsevier, Clarivate Analytics, and Digital Science. Of the around 100 million DOIs available worldwide, approximately 0.5% have an access level change every week. “With half a million publications with a DOI having a status change every week, our product has a significant impact on the accuracy of the services delivered by our customers,” says Piwowar.

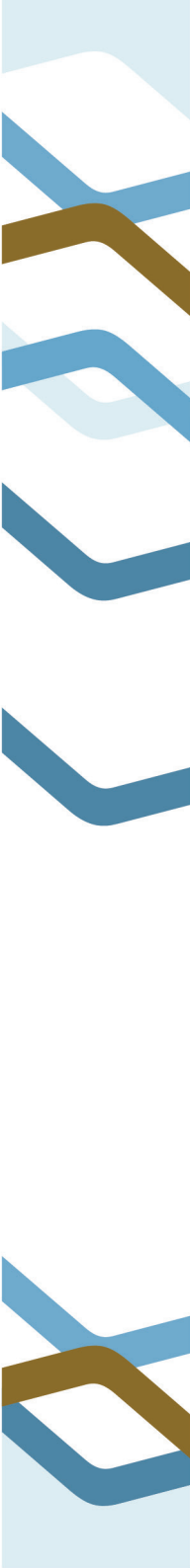
“This service brings in about half a million dollars a year and, as we don’t really have any competition there, this is a pretty steady stream of income for us which will allow us to hire a programmer to finally expand our team of two.”

Impactstory’s income sources are diverse: earned income from Unpaywall, the subcontract for research at University of Texas, and a significant grant funding stream. “Despite the success of the Unpaywall paid service we don’t plan to move away from grants, because they allow us to solve problems that the market isn’t solving,” says Piwowar. “The grants are still important for us to remain innovative, for example, right now we are building a search engine on top of Unpaywall with an **Arcadia Fund (arcadiahfund.org.uk)** grant.”

““ Despite the success of our paying service, we don’t plan to move away from grants, because they allow us to solve problems that the market isn’t solving. ””

With only the two co-founders on the payroll at the time of the interview [a third hire was made in December 2018, Gwen Franck], Impactstory is a very nimble organisation. They manage their work according to the 80/20 principle: finding out which 20 percent of the effort will solve 80 percent of the problem. Because of Impactstory’s small size it is good at changing course if something is not a good fit. These are characteristics that are more often associated with startup businesses, not with non-profit organisations, and that is not a coincidence.

““ We thought hard about whether we wanted to be a non-profit or a commercial company. I think the fact that we are so much in doubt about this, makes us different from other players in the field. ””



“We thought hard about whether we wanted to be a non-profit or a commercial company,” says Piwowar, “and I think the fact that we are so much in doubt about this makes us different from other players in the field. I don’t necessarily think that non-profits are the only ones that can be mission-driven, but we realise that our non-profit status allows us to reassure some minds, especially in academia, and it also allows us to pursue certain types of grants. And because we are not seed funded, unlike many commercial startups, we don’t need to pay back our funding.”

“ Non-profits often have a reputation for being bloated, slow, unsustainable, dependent on grant money [...] we try to incorporate the start-up mindset in our own business. ”

“Non-profits often have a reputation for being bloated, slow, unsustainable, and dependent on grant money. You become like the people you hang out most with, so that’s why we like to hang out with nimble startups in incubators! We try to incorporate the startup mindset in our own business.”

Impactstory’s small size is not the only thing that makes the organisation unique. Both Piwowar and Priem are academics, so they have in-depth knowledge of academic sensitivities, but they are also very technically oriented with a programmer mindset.

They can solve technical and coding issues themselves instead of having to hire extra staff (although, as mentioned before, they are now hiring a dedicated programmer to help out with the new project).

Both are also excellent communicators who like to do their own marketing and advocacy, and Piwowar states that they intend to continue that way: “We’ve actually never considered hiring a dedicated person for communications and marketing. Right now, we approach our communications very intuitively and I

am not sure what would be considered a good job, objectively. Despite Jason having coined the term ‘altmetrics’, we’ve learned the hard way that tweets are not a substitute for uptake. Twitter can provide a skewed perspective – it does not always reflect growth rate. Our first attempt to create a research profile product generated a lot of buzz online but this wasn’t converted into growing usage numbers.”

A downside of being a small organisation is that there’s no in-house support for administrative and legal matters and for pursuing grants, so a lot of the already limited staff time goes into working on those, instead of into developing services.



Another drawback of Impactstory's small size is that, despite being in a very good place to build innovative tools, its progress is not always as fast as desired.

"Ironically, the best-case scenario for open scholarship in general could be potentially harmful for Impactstory's business model. If every DOI were to resolve into an open access version, the Unpaywall Data Feed would be superfluous for new research. But the services would still be needed for clearing back catalogues and, as long as there are policies requiring green open access (via self-archiving), our services will definitely remain in demand."

With the number of active users continually increasing (about 1,000 new users a week), a high scoring browser extension, and a lot of positive feedback via e-mail and on social media, Impactstory is proving that the services it provides offer real added value to researchers. Because of their highly personal marketing approach, Piwowar and Priem follow up closely on interesting uses – and are always happy to get detailed feedback on how their services are being implemented.

A bigger worry is that the competition might reuse the Unpaywall code base to create their own product – which is technically possible because the code licence allows for commercial reuse: "It is a risk we run because of our liberal licensing policy. A big company might consider running it in-house, but I don't think they can do it as efficiently and cheaply as us, at least not in the US. A small company might try to offer the same service as we do. However, a competitor would still need to convince our big clients that they'd do a better job of it than us and we really have a headstart on our competitors. Changing our licence to remove this risk entirely could be an option but for now we've decided against it, mainly because non-commercial licences are notoriously difficult to enforce and non-compliance is difficult to detect."

“ Non-commercial licences are notoriously difficult to enforce and non-compliance is difficult to detect. ”

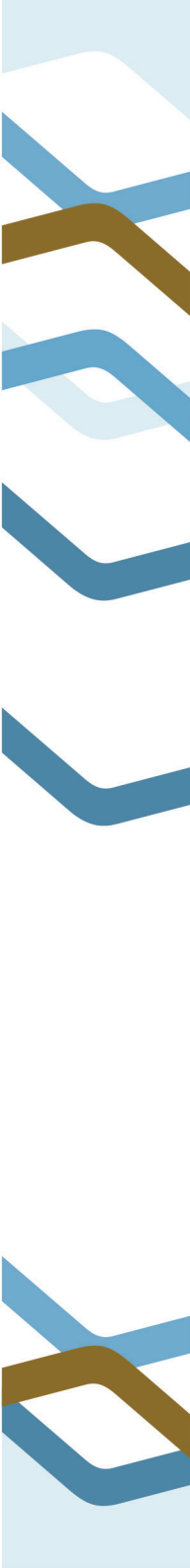
References and relevant links

- ▶ Impactstory : impactstory.org
- ▶ Sloan Foundation: sloan.org
- ▶ Duke University: duke.edu
- ▶ Depsy: depsy.org
- ▶ University of Texas at Austin collaboration: ischool.utexas.edu/tags/impactstory
- ▶ CORE: core.ac.uk
- ▶ Open Access Button: openaccessbutton.org
- ▶ OpenAIRE: openaire.eu
- ▶ BASE: base-search.net
- ▶ PubMed: ncbi.nlm.nih.gov/pubmed
- ▶ ArXiv: arxiv.org
- ▶ OAI-PMH: openarchives.org/pmh
- ▶ Crossref: crossref.org
- ▶ DOAJ: doaj.org
- ▶ DOI: doi.org
- ▶ Arcadia Fund: arcadiahfund.org.uk

About Heather Piwowar Co-founder



Dr Heather Piwowar is a co-founder of Impactstory, the non-profit company behind the Unpaywall database for open access discovery. A longtime advocate for open science, Dr Piwowar is also a leading researcher in research data availability and reuse, including a seminal paper measuring the citation benefit of publicly available research data. Dr Piwowar has a bachelor's and a master's degree from Massachusetts Institute of Technology (MIT) in electrical engineering, ten years of experience as a software engineer, and a PhD in biomedical informatics.



© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A look into Impactstory with Heather Piwowar, co-founder

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: September 13, 2018

DOI: [10.5281/zenodo.3083954](https://doi.org/10.5281/zenodo.3083954)

This report is available under a CC BY 4.0 licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info



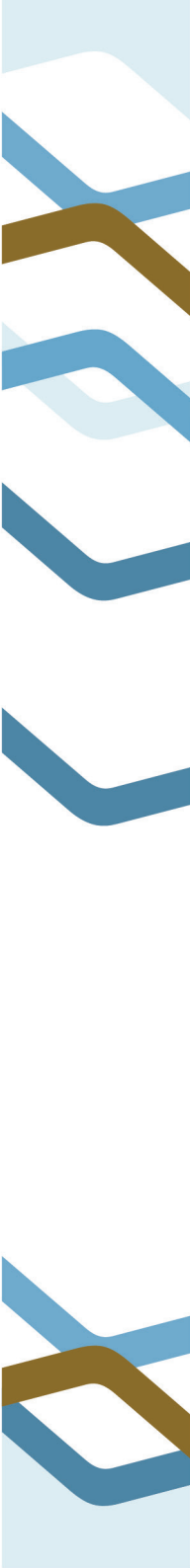
Insights into the Economy of Open Scholarship: A look into Figshare with Mark Hahnel, CEO



About Figshare

Figshare is an online digital repository where researchers can preserve and share their research outputs, including figures, datasets, images, and videos. For individual users it is free to access and to upload content. Figshare also offers a paid option licensing the infrastructure to institutions, publishers, and funders. It is one of a number of portfolio businesses supported by Digital Science.

[figshare.com](https://www.figshare.com)



Figshare: Business model

Key activities

- ▶ Repository for data and other research outputs
- ▶ Issues digital object identifier (DOI) for each uploaded item
- ▶ Licensing of infrastructure to third parties



Organisation type

- ▶ Commercial company
- ▶ Approx. 40 full-time equivalent (FTE) staff



Key partners

- ▶ Digital Science
- ▶ Institutions
- ▶ Publishers
- ▶ Funders
- ▶ Hosting infrastructure lies with Amazon



Revenue streams

- ▶ Licensing fee for infrastructure
- ▶ Part of the Digital Science portfolio



IP/Copyright

- ▶ Free version: uploaded materials CC BY or CC0 (public domain)
- ▶ Licensed (paid for) version: uploaded materials can be public domain, full copyright or all types of open licences
- ▶ Source code: closed



Customers/users

- ▶ Individual researchers
- ▶ Publishers
- ▶ Funders
- ▶ Institutions



Interview with Mark Hahnel

Being led by a researcher, Figshare has always intended to be based on researchers' needs – but with a firm nudge towards open. In the free version only very liberal open licences (or even public domain dedication) are possible, but Figshare is more flexible when licensing the infrastructure (which is a paying service).

“At Figshare, we believe that all academic outputs should be as open as possible, as closed as necessary,” says founder and CEO **Mark Hahnel**. “In the very early days we actually worked from the idea that ‘everything needs to be as open as possible’ but, as we started to operate in the global market, we developed a more nuanced view. As it turns out, not all research can and should be made openly available immediately – researchers can have bona fide reasons not to do this. But, if they want to, we are there to facilitate it.”

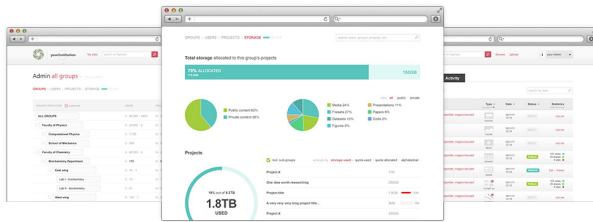
““ We believe that all academic output should be as open as possible, as closed as necessary. ””

As Figshare is operating in a commercial startup environment [currently a part of the **Digital Science** (digital-science.com) portfolio, Gwen Franck], the team has been forced from the start to think about sustainability and scalability. Moreover, states Hahnel, especially in academia, commercial activity in the field of research data management is

often frowned upon - so it is very important for Figshare to be candid about all its activities: “We have to be transparent about our workings. We are accountable to our clients and are forced to provide clarity about our budgets, our timelines and the services we provide.”

Figshare operates on two different tracks: figshare.com offers free services to end users and the company also licences the infrastructure to paying customers, such as research institutions and publishers.

The company started out as a business to consumer (B2C) service – providing free data storage services to researchers. Because Figshare only offers the option to licence the data as **Creative Commons** (creativecommons.org) Attribution (CC BY) or even put it in the public domain via Creative Commons Zero (CC0), they consider this free service as essential to their core mission.



Example of an institutional dashboard

“ Our free service only offers CC BY and CC0 as options. We believe that nudging our users in this direction – it’s ‘the price’ they pay – is essential to making research as open as possible. ”

“Our main value proposition is that we offer a free, lightweight service, not only addressing storage and facilitating sharing, but also focusing on metrics so that our users actually see the benefit of using our services. Our free service only offers CC BY and CC0 as options. We believe that nudging our users in this direction – it’s ‘the price’ they pay – is essential to making research as open as possible. With data sharing, you will always have a large proportion of users that don’t use our service because they want to make their work open, but because they have to – because they are being forced by their institution or funder,” says Hahnel. “These people will not use our services because they have made an entire cost-benefit analysis, but

because we’re convenient. They might as well use an institutional or disciplinary repository for their purpose. This realisation made us move into the business to business (B2B) model more, where we provide our services to institutions, publishers, and funders.”

The company employs over 40 full time members of staff, so a steady income is necessary to cover these overhead costs. “Licensing out our infrastructure allows us to establish a level of sustainability while keeping the researcher-oriented services free. Offering the paying service also addresses the common fear that, at some point, we will try to sell the data they accumulate with our metrics services. Because the licensing feature provides a sustainable income, this fear is unjustified,” says Hahnel.

“ Licensing out our infrastructure allows us to establish a level of sustainability while keeping the researcher-oriented services free. ”

“And about the fear of us being bought by one of the big publishers I can only say that we don’t plan to do it and, in any case, our bylaws offer enough protection for the data that we store.

“We don’t really have any alternatives to this business model. Switching our free service to an ad-based model is not realistic. We are not in the ad-selling business, we provide technology. Changing this would have a lot of logistical and legal consequences that would take up so much space.”

Figshare has offices in London, Romania and Washington DC, with remote staff working from Australia, South Africa, the US and western Europe. “If you asked us four years ago, I’d say that the general demographic profile of our staff was your typical 25-35 year-old white male, but we’re making an effort to become more diverse. With an increased focus on operations instead of technical services diversification has become necessary. If we want to expand our services to other regions we also need to ensure multi-lingual support,” says Hahnel.

There are no plans to outsource the work as these remote workers are all on the Figshare payroll. However, politico-economic trends and events such as Brexit and trade wars have made the need for local capacity clear. International travel might become more difficult in the future. But distributed working also has its issues, for example, when working with local human resource (HR) offices.

Hahnel is confident that increasing demand for Figshare services will counter any eventual logistical difficulties: “We might not offer the most competitive wages in the business, but we’re pretty lean and can offer a lot of flexibility to our employees. People want to work for us for a number of reasons but, for me, the most important thing is that they subscribe to our principles. We can provide really good infrastructure for any type of research output. The technology behind our services remains the same and we can cater to almost any request. The question is whether we want to. We don’t have the ambition to become a 2,000-person company, so we need to make tough choices now and then about whether we go after a specific part of the market. Pre-print sharing, for example – we

decided to include this in our services five years ago and that proved to be a good guess. I don’t have any set rules to decide what we will and won’t do. As CEO, I decide whether we’ll take part or not.

“In 2011, we were simply in the right place at the right time,” says Hahnel. “The technology that we rely on was there (**DataCite** (datacite.org), **Crossref** (crossref.org), **ORCID** (orcid.org), cloud computing) and allowed us to offer simple and lightweight solutions to data sharing issues researchers faced. The emergence of cloud computing, in particular, was a big help in addressing policy clashes many researchers faced – especially when it comes to data. Institutional policies, which focus on marketability and patents and the potential of spin-offs, often differ from more open-oriented funder policies.”

These policy clashes often occur: “One department, or the library, wants to implement our services but then another department sees a clash with internal intellectual property (IP) regulations. This is a major issue that needs to be addressed at supra-institutional policy and funder level.”



During the startup phase Figshare could benefit from the business savviness of its investors [at Digital Science, Gwen Franck], who, much sooner than the general research community, realised that open data policies would become a main funder requirement in the years to follow.

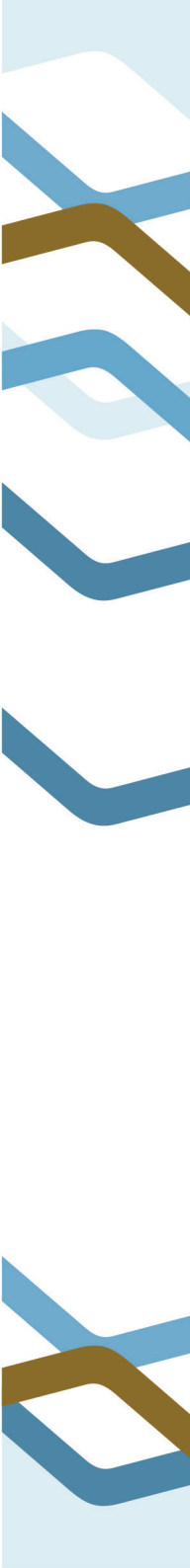
Hahnel: “A big benefit is that I am a researcher myself. I am a stem cell biologist. I had datasets, videos, and images and I had nowhere to store and preserve them while at the same time allowing easy sharing and keeping track of reuse. Especially in the startup years, I was my own guinea pig. I worked with my best interests as an academic in mind. This gave us a headstart of five years over our competition.”

“ Especially in the startup years, I was my own guinea pig. I worked with my best interests as an academic in mind. This gave us a headstart of five years over our competition. ”

But at the same time, Figshare faces a lot of criticism because of its for-profit status: “In the scientific community there is a general prejudice against commercial companies. There’s always doubt whether we will remain in business and, if we do, whether we won’t be bought by one of the big players, as happens with so many innovative startups. We can address these doubts quite convincingly, I hope. We won’t lose that many clients suddenly so that we’re driven out of business. And if we do, we’re doing something wrong and need to change the nature of what we’re offering.”

Hahnel is convinced that an advantage of being a for-profit is that it allows Figshare to be very straightforward about the services it offers, and about what customers – paying and non-paying – can expect: “In that regard, we believe we’re different from other players in the field. Our for-profit status is also very visible because we organise local events and sponsor a lot of conferences. Although the return on investment for these activities is difficult to measure it helps our people to understand the local space, to engage

with our existing users, and customers and yes, it’s also a way for us to prospect the market. We often meet resistance at these events, but it’s not because we’re in business, nor that we don’t want to understand the needs of the community. Also, I have noticed that non-commercial entities, such as universities who decide to create their own data repository from scratch, are often a lot more flexible about matters such as deadlines and accountability, leading in a lot of cases to a less efficient spending of public funds. In the name of open science they claim to offer the most ethically sound service, but in the end they cost society ten times as much as the services we provide.”



Because of its headstart in the business, Hahnel believes that Figshare will be able to offer a reliable and innovative service at a competitive price: “It’s even in our contracts. If we don’t fit anymore we help our clients move to a new system. I don’t believe in vendor lock-in and this is how we put our money where our mouth is. Despite the fact that we’re not fully open source, we do provide our clients with our code base in case we do go out of business.”

“If we don’t fit anymore, we help our clients move to a new system. I don’t believe in vendor lock-in and this is how we put our money where our mouth is.”

Hahnel is aware that that is a controversial subject: “At first sight, not being open source is not aligned with our open science principles. But we have our reasons. Not only do we want to protect our services against competition from less ethically responsible companies, we are investing a huge amount of money in staff time and technology in order to provide the best service possible to our clients.

“I don’t believe open sourcing our code would save our clients a lot of money, as a do-it-yourself (DIY) adaptation of our technology will cost them more than if they just licence it from us and we sustain it centrally,” says Hahnel. “We believe that providing a robust and scalable infrastructure as a service is a better use of public funds than having a localised service at every institution. It is perhaps not a popular opinion, but centralising services is a more efficient allocation of money – even if that means that some localised services cannot be continued and some jobs and positions will be on the line.

“We believe that providing a robust and scalable infrastructure as a service is a better use of public funds than having a localised service at every institution.”

Unlike other free data repository services, Figshare doesn’t offer much flexibility when it comes to licence choice in the free service. Users can choose between CC BY and CC0.

Hahnel: “This is of course not only in line with our open science principles – I am convinced that, if you offer people the choice, they’ll always be tempted to add a non-commercial clause – but it also has consequences for our business model. Offering true open datasets will allow far more opportunities for others to build on top of these datasets, keeping us relevant as a storage and sharing option.”

Sometimes people want to use Figshare's technology and services in a way that is not aligned with its principles. Hahnel: "We've had requests to use our platform as a tool for private data sharing – which is not difficult to do because we offer the option to make data private. These requests come from individual researchers, projects, and research groups. Universities ask us if it's possible to integrate embargo options or to make the service available only on their campuses. We get asked to install a 'request button' feature and even if it's possible to integrate a paywall! Apart from the latter request, I don't think these questions are necessarily unreasonable, although I don't think they should be a part of our services by default."

Sometimes it is legally impossible to clear all research retroactively and make it open access (for example, old theses), so Hahnel is definitely open to adding certain layers upon request: "We tread carefully there. Adding these layers is often very complicated to implement and also I am a firm believer that technology is a powerful way to nudge people into certain behaviour. When we engage with universities using our services, we

judge on a case-by-case basis. If limiting access to campus allows 50,000 people to get access to an archive of old research as opposed to no-one getting access I'd be inclined to consider it, provided, of course, that it is not applied to newly deposited research. I don't think this is contrary to our 'as open as possible, as closed as necessary' principle. We're lucky that the market has become big enough so that we don't have to engage with requests if we consider them incompatible with our principles. We can cater our services to that part of the market that explicitly wants or needs to be open."

“ We don't have to engage with requests if we consider them incompatible with our principles. We can cater our services to that part of the market that explicitly wants or needs to be open. ”

References and relevant links

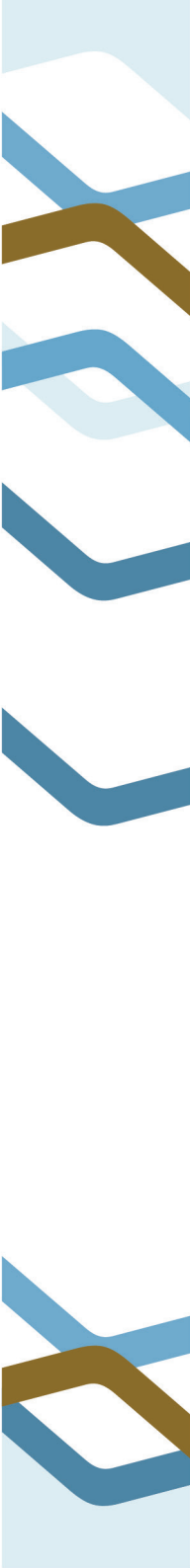
- ▶ Figshare: figshare.com
- ▶ Digital Science: digital-science.com
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ DataCite: datacite.org
- ▶ Crossref: crossref.org
- ▶ ORCID: orcid.org

About Mark Hahnel

CEO



Mark Hahnel is the CEO and founder of Figshare, which he created while completing his PhD in stem cell biology at Imperial College London.



© Knowledge Exchange 2019

Title: Insights Into the Economy of Open
Scholarship: A look into Figshare with
Mark Hahnel, CEO

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: August 14, 2018

DOI: [10.5281/zenodo.2586862](https://doi.org/10.5281/zenodo.2586862)

This report is available under a CC BY 4.0
licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info

The Zenodo logo, consisting of the word "zenodo" in a white, lowercase, sans-serif font, centered within a solid blue rectangular background.

Insights into the Economy of Open Scholarship: A look into Zenodo with Tim Smith, head of collaboration, devices and applications, CERN/IT

The logo for the Deutsche Forschungsgemeinschaft (DFG), consisting of the letters "DFG" in a bold, black, sans-serif font.The logo for Jisc, consisting of the letters "Jisc" in a bold, black, sans-serif font, with the "i" and "s" having a unique, rounded shape.The logo for the Centre National de la Recherche Scientifique (CNRS), featuring the letters "CNRS" in a black, serif font.The logo for SURF, consisting of the letters "SURF" in a bold, black, sans-serif font, with a stylized black shape resembling a surfboard or a drop to the right of the letters.The logo for the Ministry of Higher Education and Science, featuring a small crown icon above the text "Ministry of Higher Education and Science" and "Danish Agency for Science and Higher Education" in a small, black, sans-serif font.

About Zenodo

The European Organisation for Nuclear Research (French: Organisation européenne pour la recherche nucléaire), known as CERN, is a European research organisation that operates the largest particle physics laboratory in the world. Established in 1954, it is based in a north western suburb of Geneva on the Franco-Swiss border, and is funded by the governments of 22 EU member states. Zenodo is the research data repository service developed and hosted by CERN. It was created in 2013 as the OpenAIRE orphan records repository. It was relaunched as Zenodo in 2015 and allows the upload of files up to 50 GB.

zenodo.org

Zenodo: Business model

Key activities

- ▶ Repository for data and other research outputs
- ▶ Option to create communities and collections
- ▶ Issues digital object identifier (DOI) for each uploaded item



Organisation type

- ▶ Non-profit
- ▶ International governmental organisation (IGO)
- ▶ Hosted at CERN



Key partners

- ▶ CERN
- ▶ Governments funding CERN
- ▶ European Commission (via OpenAIRE)



Revenue streams

- ▶ Infrastructure and staff hosted by CERN
- ▶ Project funding (OpenAIRE)
- ▶ Experimenting with crowdfunding
- ▶ Investigating a credit model with funders



IP/Copyright

- ▶ IP is always retained by the uploader
- ▶ Materials uploaded: from public domain to full copyright and all types of open licences possible
- ▶ Own code base: open source



Customers/users

- ▶ Individual researchers
- ▶ Projects
- ▶ Civil society
- ▶ Research groups
- ▶ Conference organiser
- ▶ Institutes
- ▶ National consortia



Interview with Tim Smith

CERN was founded with the idea of international collaboration to better the world. Sharing research has always been in the CERN statutes and affiliated researchers were doing this even when it wasn't called 'open' yet, such as when they invented the worldwide web and when they started preprint servers. Recently, CERN has taken a more proactive stance, sharing the software staff have created for themselves (such as Zenodo) with the entire global research community to encourage more openness in research worldwide.

"We understand that we're part of a big research ecosystem. If the system doesn't do what we want we use it nonetheless, but using our proactive approach we try to promote better solutions," says **Tim Smith**, head of collaboration, devices and applications at CERN. "We produce fundamental research that benefits the whole world: it's pure science. Open scholarship for us means that we enable others to use the same tools and techniques as we do, to follow our research or to use it in their own so that everyone benefits from our research innovations as well as our research outputs."

"Zenodo fits this ideal perfectly – we built the underlying technologies for ourselves, and then used input from other disciplines and experts to share it as a service for the entire research community."

“ Open scholarship for us means that we enable others to use the same tools and techniques as we do, to follow our research or to use it in their own so that everyone benefits from our research innovations as well as our research outputs. ”

"Over the last decades, operating at the ever-expanding big data frontier has led us to new research workflows and to set up services drawing together computer centres across the world in a world-wide data grid," says Smith. "Opening up these datasets changes the dynamic once again. Accessible and reusable data requires even more resources than we anticipated 20 years ago, when we wrote the proposals for the projects we are still working on. With small data there's no issue, as the effort and cost is marginal. With our larger datasets, however, we have to take into account that nobody else has the tools to work with them, let alone reproduce the research."

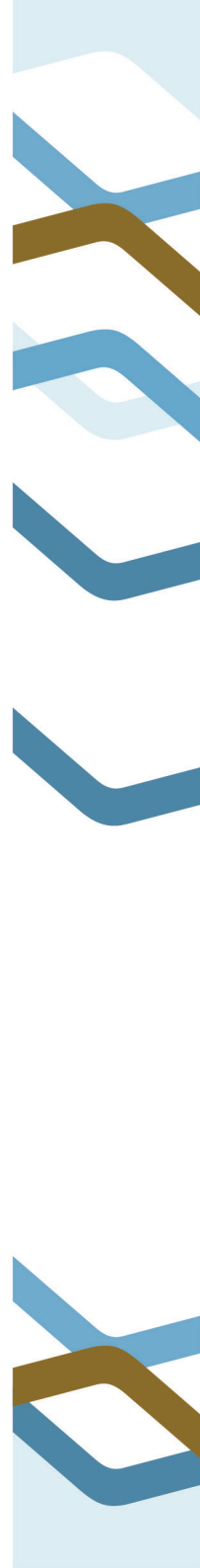
“Overcoming these obstacles is something that we are actively looking into right now. Opening up and managing massive datasets is a learning process. What is already clear is that you need to capture relevant information as soon as you can while the research is ongoing, including the tools and metadata and intermediate datasets, otherwise reproduction of the research might become impossible afterwards.”

“Due to the scale, a lot of our research is not practically reproducible by others, so we need to duplicate and validate our research ourselves... this produces a very interesting internal dynamic.”

CERN considers itself capable of tackling most technical challenges. “The human aspect, however, is something completely different; changing existing research workflows such as publishing practices requires a lot of effort,” says Smith. “At CERN, we are in a unique position with a hand in the entire research ecosystem. Due to the scale, a lot of our research is not practically reproducible by others, so we need to duplicate and validate our research ourselves by having independent apparatuses and analysis chains. This produces a very interesting internal dynamic.”

“For us, it is necessary to have a ‘closed’ research phase wherein each experiment prepares its findings independently. Our tools therefore support the entire spectrum of research activities and outputs: from datasets that are temporarily closed, to research that has been finalised and is ready to be shared with the world.”

Zenodo was created to offer other sciences the same capability to capture all research in all phases – regardless of its closed/open status. But Zenodo is also aiming for something more aspirational. Smith: “With Zenodo we can show everybody not only what can be done, but also what could and should be done when it comes to research data management. Outside of CERN all researchers can rely on this service, provided their needs are not too great. It is feasible for us, both because we want to do this for the greater good of open science and also because, as our overall infrastructure is so massive, it is a marginal activity that only requires a small addition to our resources.”





Display of uploaded research article on Zenodo

At the moment, the operation of Zenodo is a peripheral activity for CERN, as the size and quantity of the datasets stored in Zenodo is easily manageable compared to its own big data. As long as Zenodo's growth remains less than the growth of CERN's own datasets, it will remain like that. Due to this unique situation, there is no urgency for a self-sustaining business model attached, as the organisation feels capable of supporting the current needs of the community without too many additional resources.

The biggest risk however seems to be that the service will become the victim of its own popularity. "If the rate of dataset growth outpaces CERN's natural data centre evolution, we'll have to incorporate this in our cash flows as it will no longer be a marginal activity. It's technically possible, but it would require a more business-oriented approach between funders, researchers, and Zenodo," says Smith. "Moreover, it needs to be taken into account that data is not static. Versioning is easy with software but applying the same techniques to mega datasets will rapidly become unmanageable if we don't keep pace with new techniques."

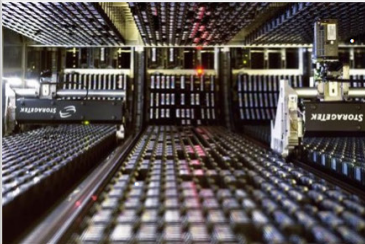
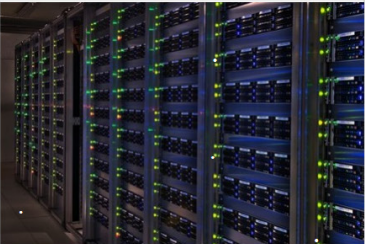

Although there's currently no formal business plan, CERN is definitely investigating potential new funding streams in order to anticipate Zenodo's sustainability.

“ Money needs to be earmarked for data services regardless, so why not invest it in a common infrastructure such as Zenodo? ”

Smith: “We are investigating a credit model with major research funders. As CERN is an intergovernmentally funded organisation, one way of doing this sustainably would be to transfer funds from one national infrastructure to another via a credit system, retroactively according to usage. Money needs to be earmarked for data services regardless, so why not invest it in common infrastructures such as Zenodo? In my opinion, this

would be the most efficient way to apportion existing resources to long-lasting services. But I feel funders and research-performing organisations are not quite ready for it – they see it needs many changes to cash flow processes. Our vision of shared infrastructures, common tools, and solutions requires a seachange in how funders, states, and other institutions spend their resources.”

Investigating a different tack, CERN is also experimenting with using microfunding. They launched a **donation site** (<https://bit.ly/2PC3urO>) for Zenodo users and supporters, being completely transparent about how much it costs to run the service. “The uptake is very low though, not least because there are a lot of roadblocks in international research funding. For example, project funding cannot always be awarded for non-contract or non-commercial services. We are currently experimenting with a lot of different funding models, but our principal aim is to try and contribute to a standard shared infrastructure for managing research data. The long-term custodianship of resources that benefit the world should, in my opinion, be governed as a commons.”

| Big Data Founded | Cloud Connected | Human Powered |
|--|--|---|
|  |  |  |
| <p>220 CHF 1TB in Zenodo for 10 years</p> | <p>500 CHF One day of new feature developer</p> | <p>2500 CHF One week of overall service management</p> |

Crowdfunding for Zenodo - donation site (<https://bit.ly/2PC3urO>)

At this point, Smith deliberately refuses to calculate how much working on open science costs CERN in terms of staffing: “It is a misleading number without a lot of context. Zenodo isn’t run as a standalone service; it relies on the multi-level cloud services we have created in our data centre for server provisioning, service orchestration, storage, databases, and networking, which are already staffed for a very high capacity and complexity of workload.”

“It also relies on the research data management programmes we have written for our own open science services. Funding via European Commission (EC) projects such as **OpenAIRE** (openaire.eu) provides the necessary extra staff to operate and support the Zenodo service on top. In addition to the grant value, we could then add a fraction of the cost of each cloud service layer and a slice of the digital library development effort. This would then represent the cost of growing a marginal service, but not the cost of creating another at CERN or elsewhere, nor the cost of running a larger scale operation. And it doesn’t take account of the fact that we would maintain the same staffing level and development effort anyway!”

While CERN is not the only organisation or company offering research data management services, Smith sees a big difference from their competitors: “We are often represented as one of several similar options, but we are completely different from our colleagues at for-profit companies, the main difference being trust. Trust is a rare and hard-won currency these days and we hope that open science helps to re-establish people’s trust in science in general. That’s why you need organisations that are worthy of such trust, like CERN, to build shared services.”

An essential element to establish this trust is, in Smith’s opinion, the non-profit status of the organisation: “Not every commercial company is dishonest, it’s simply that if you’re driven by growth requirements you’ll either monopolise the market or you’ll get acquired by bigger players, that’s the scheme of business.”

“Therefore, I think the larger commercial operators can spawn innovation in the short term and provide valuable transient services, but in the long term they don’t serve the aims of open science. Don’t forget that in tech and open science even the medium term is very, very short.”

“Evidence shows that no matter what guarantees the commercial companies sign up to, the moment they change leadership or get taken over they’re no longer capable of fulfilling those promises.

“For-profits can be involved in parts of the research lifecycle, but they should never be allowed to own any part of it. They can facilitate, but they cannot be the arbiter or gatekeeper.”

“Commercial companies, especially startups, are very good at turning an idea into a real service. They are much better than research organisations at marketing and fundraising, and hence in service innovation, but in the consolidation phase they are often swallowed and their innovative aspect gets lost,” says Smith. “That’s why for-profits can usefully supply parts of the research lifecycle, but they should never be allowed to own any part of it. They can facilitate, but they cannot be the arbiter or gatekeeper, for instance, when it comes to metrics.”

Contrary to competing services such as **Figshare** (figshare.com), which in its free service only offers the most liberal **Creative Commons** (<https://creativecommons.org>) licences (CC BY and CC0) as options, Zenodo allows the full set of licences including those with more restrictive clauses. Smith: “We want to enable everybody to participate in open science. This does not mean that we consider all licences as equal. But we cannot foresee what will happen in every field at any given time in the future, not even in our own field. Therefore, we think the best, the easiest, and the most persuasive service

we can give is to offer researchers what they think they need, and to gently nudge them in the direction of open science best practices in the long run.”

““ We think the best, the easiest and the most persuasive service we can give is to offer researchers what they think they need, and to gently nudge them in the direction of open science best practices in the long run. ””

Zenodo is open source, based on **Invenio** (<https://invenio-software.org>) software, which is a digital library framework CERN invented in the nineties. “Naturally the world has caught up on many fronts and there are now alternative components available, so we need to be flexible and agile and replace and incorporate external pieces into our framework,” says Smith. “When considering externally developed

products, we always look at open source solutions first. For example, we replaced our own search engine with **Elasticsearch** (elastic.co). We do this because we believe choosing open source is our best way of guaranteeing independence and longevity, and of increasing collaboration and promoting inclusion. But we are pragmatic at CERN, hence we operate with the best, most economical, and functional tools we can find.

“Sometimes, especially when it comes to hardware, we need to go to the biggest commercial suppliers because only they can offer the tools we need. For software, we have more flexibility and we are able to bring our principles into practice more. I use an informal decision tree that allows me to identify suitable open source tools and whether the cost of running, adapting, and maintaining them is justified. “CERN doesn’t need Zenodo in order to continue its activities and Zenodo has never been intended to run at a profit. Our aim is to showcase the possibilities, to facilitate, and stimulate open science practices that were considered to be nearly impossible by most people,” explains Smith.

“Ultimately Zenodo is a means to share with the world what we have learned and developed and contribute to a common and shared infrastructure for all. If somebody comes up with a better way to run such services, with equally good or better guarantees of openness, I would be happy to pass on the baton!”

“ CERN doesn’t need Zenodo in order to continue its activities and Zenodo has never been intended to run at a profit. Our aim is to showcase the possibilities, to facilitate, and stimulate open science practices that were considered to be nearly impossible by most people. ”

References and relevant links

- ▶ CERN: home.cern
- ▶ Zenodo: zenodo.org
- ▶ Zenodo donation page: zenodo.org/donate
- ▶ Creative Commons licence suite: creativecommons.org
- ▶ Figshare: figshare.com
- ▶ Invenio: invenio-software.org
- ▶ Elasticsearch: elastic.co

About Tim Smith

Head of collaboration, devices and applications, CERN/IT



Tim is an open science advocate leading initiatives at CERN and in the wider science community. He drove the launch of CERN Open Data, a portal to share Large Hadron Collider (LHC) big data with the world, as well as the Higgs boson webcast, which shared its discovery live around the globe. He also instigated and nurtures Zenodo within the European Commission's OpenAIRE project as an open data service for worldwide science. Tim came to CERN at the end of the 80s, obtained a PhD in particle physics and performed research at the Large Electron-Positron Collider (LEP) for ten years. He then joined the CERN IT department to lead teams innovating in computing farm management and physics data management.

© Knowledge Exchange 2019

Title: Insights into the Economy of Open Scholarship: A look into Zenodo with Tim Smith, head of collaboration, devices and applications, CERN/IT

Authored by: Gwen Franck

Contact: gwenfranckgcv@gmail.com

Interview date: August 14, 2018

DOI: [10.5281/zenodo.2842593](https://doi.org/10.5281/zenodo.2842593)

This report is available under a CC BY 4.0 licence creativecommons.org/licenses/by/4.0



Knowledge Exchange Office
C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info

Current state and approaches for further exploration

When asked about how they position themselves in the scholarly ecosystem, the interviewees can be divided into three groups. The first group is trying to solve a 'legacy issue' in the scholarly communications market such as ScienceOpen (moving beyond the scientific journal), HRČAK (digitisation), and ASAPBio (the issue with preprints and peer review). The second group is trying to find an answer to issues that have only arisen with the rise of Open Science such as OLH (funding for open access in HSS), HUP (funding for open access monographs), Figshare and Zenodo (RDM-related issues) and Impactstory (status tracking). The third group consists of small-scale activities such as Opatnet and deals with the facilitation of collaboration at micro-level.

Despite these different positions, all interviewees agree that the current state of Open Scholarship is far from perfect and they have identified approaches needed to establish a full and functioning Open Scholarship market (or at least, to remove the gaps that prevent them from playing their role in it). In the light of Knowledge Exchange's recent work designing an **Open Scholarship framework**

([ji.sc/31LpDtI](https://doi.org/10.1017/ji.sc/31LpDtI)) and the book 'Open Scholarship and the Need for Collective Action' (will be released in autumn 2019), these gaps and the context given may indicate areas and issues that are worth further exploration.

1. Approaches to reach a critical mass

Open Scholarship initiatives should not only serve researchers who take an active interest in working in the open. Usability, flexibility, and a willingness to invest time in marketing/promotion are core factors in order to appeal to a broad mass of users.

We can identify a number of common factors that prevent a service provider from attracting enough users for a service to achieve sustainability. As identified by our interviewees, these impediments include lack of awareness about the service, the availability of better-known other (commercial) services or products, lack of satisfaction with the products and services provided (flexibility and usability) and lack of interest in open workflows in general. With current developments in Open Scholarship policy-making, awareness about the topic seems to have increased somewhat but what

is needed is more coordination at the meso-level. In order to reach a critical mass of users or customers for independent service providers such as the ones we interviewed, open workflows need to be firmly embedded in the research lifecycle – something that will not happen if relying on the actions of individual researchers alone.

Suggested approaches:

- ▶ Support for qualitative training on Open Scholarship-related topics and professional advice on available services and tools other than those offered by the big commercial players (who have an accompanying marketing budget and are often entrenched at institutional level)
- ▶ Continuing availability of grant support for small businesses in order to improve usability of their services

2. Approaches to change the licensing culture

There is definitely a need for more support on open business models and how open licensing can help with that. Often, intellectual property (IP) support offered at institutional level is patent-focused, leaving it very difficult for 'open' initiatives to get proper advice on alternative IP schemes. Although there is not much debate amongst our interviewees about the value of CC licences, most of them indicate that the consequences of applying licences are often not entirely clear for researchers. Local Creative Commons chapters can offer advice on this, but in order to obtain legally sound advice, legal services at research performing organisations and funders need to be aware of them.

While all of our interviewees show a preference (enforced or not) for 'liberal' open licences such as CC BY, it is recognised that this does not necessarily reflect the preferences of the average researcher. Many researchers believe that omitting non-commercial or no-derivatives clauses from their open licence makes them vulnerable to plagiarism and the loss of academic credit. If the aim is to

convince researchers about the value of the more liberal licences, these issues need to be addressed. A subject that all of our interviewees agreed upon, however, is that non-commercial clauses are in fact useless to protect assets – mainly because they are very difficult to enforce (not in the least because commercial use is not straightforwardly defined).

Suggested approaches:

- ▶ As the finer elements of open licensing can remain a controversial subject, even for licensing specialists, there is more action needed here than simply providing adequate training for individual researchers – although this remains a crucial factor, especially with regards to credit and citation. What needs to be addressed as well is the often incorrect application of licences by research performing organisations and service providers (inadequate or even absent licence policies), and by publishers (for example, the 'copyrighting' of non-copyrightable materials such as metadata)

- ▶ Clarity about licensing requirements needs to be embedded in all Open Scholarship policies at all levels, and the input for this needs to be provided by legal professionals who have a thorough knowledge about open licences

3. Approaches to avoid data and vendor lock-in

Traditionally, a couple of major 'heritage' players have been shaping what is recognised as 'research output' – strongly tied to the business models of these players. If we want to make sure other actors and research products are also recognised as part of the 'story' of research, we will have to make sure that we avoid monopolies of a few major players and instead stimulate competition from more diverse providers/platforms.

As has been mentioned in the introduction, none of the revenue streams generated by our interviewees are based on other value propositions than the sale of data or research objects. From a vendor perspective, all interviewees for whom it is relevant claim that they do not lock in any data and that they make it as easy as possible for their customers to

move away, for example, in the case of an external acquisition or if the services provided do not meet customers' needs anymore. Some make this explicit in their contracts, others do not have explicit policies in place and claim to work on a case-by-case basis. In some cases, interviewees (working in a larger institutional context) faced with make-or-buy decisions have voiced some frustration with overarching institutional policies that often don't provide sufficient measures against vendor lock-in.

Suggested approaches:

- ▶ Do not accept contracts that contain clauses establishing vendor or data lock-in
- ▶ From the vendor side: include clauses that ensure that no data or vendor lock-in will happen. Have sufficient and detailed measures in place and include them in any contracts signed

4. Approaches for incentivising innovation

As has been pointed out by several interviewees, libraries especially are facing budget cuts. Scholarly communication (and especially publishing) is a sector where a lot of money is invested in order to obtain access, often via package deals. It would be sensible to invest part of that money in initiatives that provide truly open access to materials.

It is striking that innovations such as preprint publishing and open peer review are mentioned by quite a few interviewees as things that they wanted to work on or implement at a very early stage, but they had to be shelved because they did not gain any traction. Whether a tool or workflow gets picked up is not only related to policy changes and technological advancements, but also to changing demand amongst researchers.

Innovation, however, is not limited to services and products but also relates to business models. A common fear is that, even in an evolving area such as the Open Scholarship landscape, there will be a convergence into one or two 'established' models,

disregarding all alternatives (an example often mentioned is the APC-based model for open access publishing). One of the aims of this report was to showcase other business models and offer some examples of alternative approaches.

Suggested approaches:

- ▶ Although not all alternative business models will turn out to be solid or scalable, attention to these models needs to increase. Especially at policy level, certain ‘alternative’ models can prove to be a better application of public funds
- ▶ When it comes to innovative tools and workflows, their use must not hinder a researcher’s career progress. Therefore, their recognition in research assessment is crucial – ultimately leading to a change in demand towards more innovative approaches from the end user side



Appendix 1: methodology

Participants

The interviews have been selected in order to represent a broad overview of the different types of actors that provide Open Scholarship services. Taking into account a certain level of geographic distribution (although this was not a decisive factor), the Knowledge Exchange Task & Finish Group on 'Insights into the Economy of Open Scholarship - A collection of interviews' selected ten interviewees:

1. Open Library of Humanities (OLH) (UK): Martin Paul Eve
2. OpenEdition (FR): Pierre Mounier
3. Opasnet (FI): Jouni Tuomisto
4. ASAPBio (US): Jessica Polka
5. ScienceOpen (GER): Stephanie Dawson
6. HRČAK (HR): Jadranka Stojanovski
7. Helsinki University Press (HUP) (FI): Leena Kaakinen
8. Impactstory (US): Heather Piwowar
9. Figshare (UK): Mark Hahnel
10. Zenodo (CH): Tim Smith

Although each organisation's size and activity type varies, they can all be located in the economic

arena of the **KE Open Scholarship framework** ([ji.sc/2HwocpF](https://doi.org/10.1017/9781009054444)). Moreover, they are actors at the micro (individual or research group) or meso (journals, publishers, and repositories) levels. We did not interview actors at the macro level. The actors cover different research phases.

Interview questions

1. What do you see as the main function/ requirement of Open Scholarship that you and/ or your business or organisations contribute to?
2. What is Open Scholarship?
3. Can you describe how your initiative tries to influence and support the traditional research workflow moving in the direction of Open Scholarship? What were the target initial needs/ ambitions of your audience? What are the hoped for and expected changes in existing research workflows that your initiative wants to accomplish?
4. Have you been successful in achieving your ambitions – and can you elaborate on the reasons why you did or did not?
5. To ask about the process: what steps did you take, and why? What nudged you or your

organisation in one direction or the other during the process?

6. What is the intellectual property (IP) status of your own codebase, outputs and/or your aggregated content? How do you make sourcing decisions for your own hardware and software? Does this influence your business model?
7. What is your opinion on the role of commercial actors providing services for Open Scholarship? What should be the rules of engagement when working with commercial providers?
8. Can you provide a SWOT (strengths, weaknesses, opportunities, threats) analysis of your own organisation?

Procedure

The interviews have been conducted by Gwen Franck commissioned by Knowledge Exchange. The majority of the interviews were conducted online using Skype or a Zoom meeting room. The interviews have been recorded using the 'easy voice recorder' app. Afterwards the interviews have been transcribed and written into a story format.

Appendix 2: glossary

| | | |
|----------------|--|---|
| APC/BPC | Article/book processing charge: an author fee certain publishers charge to publish an open access article or book. | |
| API | An application program interface (API) is a set of routines, protocols, and tools for building software applications. In general terms, it is a set of clearly defined methods of communication among various components. A good API makes it easier to develop a computer program by providing all the building blocks, which are then put together by the programmer. | en.wikipedia.org/wiki/ Application_ programming_interface |
| arXiv | A repository of electronic preprints approved for publication after moderation. It consists of scientific papers in the fields of mathematics, physics, astronomy, electrical engineering, computer science, quantitative biology, statistics and mathematical finance, which can be accessed online. | arxiv.org |
| BASE | Search engine especially for academic web resources. BASE is operated by Bielefeld University Library. | base-search.net |
| COAR | The Confederation of Open Access Repositories (COAR) is an international association with over 140 members and partners from around the world representing libraries, universities, research institutions, government funders and others. COAR brings together the repository community and major repository networks in order to build capacity, align policies and practices and act as a global voice for the repository community. | coar-repositories.org |

| | | |
|--|--|--|
| CORE | CORE aggregates open access research outputs from repositories and journals. It harvests research papers from data providers from all over the world including institutional and subject repositories as well as open access and hybrid journal publishers and makes them available to the public. | core.ac.uk |
| Creative Commons | Creative Commons (CC) is a global nonprofit organisation that enables sharing and reuse of creativity and knowledge through the provision of free legal tools. The organisation has released several copyright licences, known as Creative Commons licences. | creativecommons.org |
| CrossRef | Not-for-profit association dedicated to facilitating persistent cross-publisher citation, linking in online academic journals. | crossref.org |
| Data Management Plan (DMP) | A DMP is a formal document that outlines how data is to be handled both during a research project and after the project is completed. | |
| DataCite | Not-for-profit organisation which aims to improve data citation. | datacite.org |
| Digital Rights Management (DRM) | Access-control technology (via hardware or software) restricting the unauthorised use of proprietary works. | |
| Digital Science | Company focused on strategic investments into start-up companies that support the research lifecycle. | digital-science.com |
| DOAJ | The Directory of Open Access Journals (DOAJ) is an online directory that indexes and provides access to quality open access, peer-reviewed journals. | doaj.org |
| DOI | A Digital Object Identifier (DOI) provides an actionable, interoperable, persistent link to a digital object. | doi.org |

| | | |
|--------------------------|--|---|
| Elasticsearch | Open source distributed search and analytics engine created by Elastic. It helps users to look for data safely and reliably, and to analyse and visualise data. | elastic.co/products/elasticsearch |
| Embargo | Period during which a subscription-based journal article needs to remain behind the paywall. After this period, the article can become open access (via self-archiving). Embargo terms accepted by funders vary between six months for STEM and twelve months for HSS. | |
| FAIR | FAIR data is data which is findable, accessible, interoperable and reusable. | force11.org/group/fairgroup/fairprinciples |
| GNU | The GNU General Public License is a free, copyleft licence for software and other kinds of works. | gnu.org/licenses/gpl-3.0.en.html#content |
| Gold open access | Gold open access is a term used to define open access publishing (ie the publisher makes all articles and related content available for free on its website). | |
| Google Scholar | Web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines. | scholar.google.com |
| Green open access | Green open access is a term used to define all forms of providing open access via self-archiving in a repository. | |
| HSS (also: SSH) | Abbreviation for humanities and social sciences. | |
| Hybrid journal | Journal offering fee-based (APC) open access to certain articles, but remaining subscription-based as a whole. | |
| Mega-journal | Cross-disciplinary, peer reviewed (open access) journal, accepting all articles that pass a certain quality threshold. | |

| | | |
|----------------------------------|---|--|
| Metadata2020 | Collaboration that advocates for richer, connected and reusable open metadata for all research outputs, which will advance scholarly pursuits for the benefit of society. | www.metadata2020.org |
| MIT License | The MIT License is a permissive free software licence originating at the Massachusetts Institute of Technology (MIT). As a permissive licence it puts only very limited restriction on reuse and has, therefore, an excellent licence compatibility. | opensource.org/licenses/MIT |
| OAI-PMH | The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability. Data providers are repositories that expose structured metadata via OAI-PMH. Service providers then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of six verbs or services that are invoked within HTTP. | openarchives.org/pmh |
| OASPA | OASPA represents the community of open access scholarly publishers and related organisations. It is committed to developing and disseminating solutions that advance open access, preserve the integrity of scholarship and promote best practice. | oaspa.org |
| OJS | Open Journal Systems (OJS) is an open-source software for the management of peer-reviewed academic journals, and is created by the Public Knowledge Project. | openjournalsystems.com |
| Open Access Button | The Open Access Button is a browser bookmarklet which registers when people hit a paywall to an academic article and cannot access it. It tries to link people to free, legal, full text articles instantly. | openaccessbutton.org |
| Open Knowledge Foundation | A global non-profit organisation focused on realising open data's value to society by helping civil society groups access and use data to take action on social problems. | okfn.org |
| Open source | Software for which the original source code is made freely available and may be redistributed and modified. | |

| | | |
|--------------------|--|---|
| OpenAIRE | A European project supporting Open Science. On the one hand, OpenAIRE is a network of dedicated Open Science experts promoting and providing training on Open Science. On the other hand, OpenAIRE is a technical infrastructure harvesting research output from connected data providers. | openaire.eu |
| OPERAS | A European research infrastructure for the development of open scholarly communication in HSS. | operas.hypotheses.org |
| ORCID | The ORCID iD is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. | orcid.org |
| Paywall | Term used to describe the status of non-open access articles that are 'behind a paywall' if a (subscription) charge needs to be paid to access it. | |
| Peer review | Peer review is the evaluation of scientific, academic or professional work by other scholars or researchers working in the same field (peers). Traditionally, peer review happens anonymously, but new forms of more open peer review processes, where the names of reviewers are given, are gaining traction. | |
| Plan S | Initiative for open access publishing that was launched in September 2018. The plan is supported by cOAlition S, an international consortium of research funders. Plan S requires that, from 2021, scientific publications that result from research funded by public grants must be published in compliant open access journals or platforms. | coalition-s.org |
| PLOS | Public Library of Science (PLOS) is a nonprofit open access science, technology, and medicine publisher, innovator, and advocacy organisation with a library of open access journals and other scientific literature under an open content licence. | plos.org |

| | | |
|-------------------|---|---|
| Postprint | Version of an article after peer review but before it has been laid out by the publisher. For non-open access articles, often this is the version that the publisher allows to be self-archived. Also called the author accepted manuscript (AAM). | |
| Preprint | Version of an article before peer review. | |
| PubMed | A free digital archive for citations and/or fulltexts that is developed and maintained by the National Center for Biotechnology Information (NCBI) at the US National Library of Medicine (NLM) located at the National Institutes of Health (NIH). PubMed comprises more than 29 million citations for biomedical literature. | ncbi.nlm.nih.gov/pubmed |
| R software | R is a free software environment for statistical computing and graphics. | r-project.org |
| RDF | A Resource Description Framework (RDF) is a framework for describing resources on the web. It is designed to be read and understood by computers. RDF is not designed for being displayed to people. It is written in XML. | w3.org/RDF |
| Repository | A repository is a digital archive for research outputs. Institutional repositories archive research outputs from researchers affiliated with these institutions. Disciplinary repositories accept all outputs within a certain field regardless of where the author is based. Depending on the policies in place and the infrastructure used, repositories can accept different types and versions of research outputs. | |
| RDM | Research Data Management (RDM) includes all actions needed to make research data discoverable, accessible and understandable in the long term: organisation, documentation, storage, sharing, and archiving. | |
| SciELO | Scientific Electronic Library Online (SciELO) is a Brazil-based bibliographic database, digital library and cooperative electronic publishing model of open access journals. | scielo.org |

| | | |
|------------------------|---|---|
| Scopus | The largest abstract and citation database of peer-reviewed literature including scientific journals, books and conference proceedings. Scopus is owned and hosted by Elsevier BV. | scopus.com |
| STEM | Abbreviation for science, technology, engineering, and mathematics. | |
| STM Association | Global trade association of science, technology and medicine publishers. STM members include learned societies, university presses, both subscription and open access publishers, new starts and established players. | stm-assoc.org |
| Ubiquity Press | UK-based multidisciplinary open access publisher for journals, books, and data. | www.ubiquitypress.com |
| Web of Science | Web of Science (previously known as Web of Knowledge) is an online subscription-based scientific citation indexing service originally produced by the Institute for Scientific Information (ISI), later maintained by Clarivate Analytics (previously the Intellectual Property and Science business of Thomson Reuters). It gives access to multiple databases that reference cross-disciplinary research, which allows for in-depth exploration of specialised subfields within an academic or scientific discipline. | webofknowledge.com |
| Wiki | A wiki is a website or database developed collaboratively by a community of users, allowing any user to add and edit content. | |
| Wikimedia | The Wikimedia Foundation, Inc. (WMF, or simply Wikimedia) is an American non-profit and charitable organisation. The foundation was founded in 2003 by Jimmy Wales as a way to fund Wikipedia and its sibling projects through non-profit means. | wikimediafoundation.org |
| XML | XML (eXtensible Markup Language) is a software- and hardware-independent tool for storing and transporting data. | w3schools.com/xml/xml_what_is.asp |

Knowledge Exchange Office

C/ O Jisc,
One Castlepark,
Tower Hill,
Bristol, BS2 0JA

t: 0203 697 5804

e: office@knowledge-exchange.info