Open Library of Humanities
Martin Paul Eve (University of Lincoln)

I am a lecturer in English Literature at the University of Lincoln in the United Kingdom and am on several national level panels for the UK government relating to Open Access in the United Kingdom, particularly in regards to humanities and monographs. I am also starting a project called the Open Library of Humanities that I am going to talk to you about today.

My talk today is going to be broken down into four sections. The first will go into the background of Open Access in which I will set out a series of problems we have to overcome. The second will address the social problems of Open Access. The third will be about technological solutions, and the fourth will be about financial problems and solutions. There is a famous saying by a man called Jim Gray, “May all your problems be technical, because those are the ones we can solve.”

I am going to talk about the problems as related to my project, the Open Library of Humanities. I also want to think about what we are doing and how it fits in to these various problems (#2). I will outline these problems in detail all the while thinking of Open Access as the tip of the iceberg. I will start with drivers for Open Access and its historical background. This graph shows the cost of providing access to journals in red against a measure of inflation called the UK CPI. According to the American Research Library’s statistics, since 1986 there is a 300% difference between these figures (#5). Inflation has risen by 80% whereas the cost of providing our researchers with access has risen by 380%. This is not a sustainable business model.

The subscription model is not sustainable and is economically problematic. We know also that it creates problems for researchers of not having access to material. The general public is unable to see material they have theoretically paid for through tax payer funded research. What are the alternatives? In the UK we have an organization called Research Councils UK that conducted a study into how Open Access should be implemented. They concluded that they had a preference for the gold model based on the Article Processing Charge (APC) system where authors and institutions pay up front for publishing (#6). They proposed to implement this by giving a block grant to each institution in the UK proportionate to their value judgments on the institution’s research output. Smaller institutions receive a tiny amount of funding, while big research institutions like Oxford and Cambridge receive substantially more. This is to be used to pay publishers for the services of making articles available for free.

The problem is that researchers need to publish in certain venues that hold prestige. If those venues switch to a model that requires the author or institution to pay upfront, but their institution is not given a substantially large amounts of money to do so, there is almost always a situation where there are not enough funds across the institution for everyone to publish. We then get a model of exclusion and that is deeply troubling to many academics in the humanities who work on an unfunded basis. The final economic point I want to make is that, at present, a for-profit model for publishing appears to be causing us damage. Elsevier, for instance, made £724 million in profit in 2011. At the same time, many research institutions were unable to get access to materials needed for their research. It is a social ill when researches are carrying out a public good and that needs to be fixed. This is the biggest problem we have to target.

One of the reasons I think we have this problem is what I call the "Prestige Trap." Historically, a subscription based journal could attract good submissions and accumulated prestige. We know that journals facilitate quality, and we want to continue with that model to some degree. However, I want to argue that, in the present climate, there is a real problem with evaluating at the journal level. Journals thereby extort us, because they use their reputation at the journal level rather than at the level of the authors. This forces many to continue publishing in known venues. The Prestige Trap is a curious thing because it is academics that confer prestige. They write the articles, they review the articles, and they edit the journals, so it seems strange that we are locked into an economically disadvantageous commercial system that closes off research. Simultaneously, what we need to do when building new models is to emulate the old models of prestige so we can get around this problem. You cannot change an entire system overnight. As an example, at a recent job hiring committee, we had 350 applicants for a single post. There was no way we could evaluate their research other than by looking at the name of the journal in which they had published. Prestigious journals are being used as a proxy for something that cannot really be reduced and does not capture detail.

The final of these "tip of the iceberg" points are implicit standards of peer review. Open access does not mean that we have to change our models of peer review. We can do things exactly as they stand. However, it does provide us an opportunity to re-think what we are doing. The humanities operates almost universally on a double-blind model of peer review. We can change that to models of post review or peer-to-peer review in the future.

I would like talk about how I have begun to address these problems through the Open Library of Humanities project

(#9). Regarding prestige, it useful to think about who bestows authority in the academy. It is university academics, and to make a venue acceptable for younger career and mid-career academics to publishing, we have to ensure that it is safe from the very top. It is for that purpose that we have recruited an international high profile steering committee to guide us. We have David Palumbo-Liu of Stanford Comparative Literature, we have Kathleen Fitzpatrick, chair of the MLA Scholarly Communications Department, we have the Chair of History at Harvard, David Armitage, and we have Michael Eisen, one of the founders of PLOS, the Public Library of Science, which is a major open access initiative in the natural sciences. We have also been working on an internationalization committee, as one of the problems with our committee is that it is too Anglo-American centric. This is mainly due to linguistic barriers. If you are interested in helping us with these initiatives, please get in touch. Regarding this internalization front, one of the aims of the project is for it be trans-linguistic; it is to try and overcome the mentality that we often have in the West of being knowledge exporters. This idea is false, especially when we talk about cultural dialogues. I want to break that down, and we can only do that with international participation.

You might think that starting a new initiative with all the right people in place would guarantee that there is a quality peer review procedure would make it easy to get academics to submit to you. Unfortunately, that is not the case in any way. As we were a small scholar-run initiative, we do not have a marketing budget, and we are competing with global multinational entities that are threatened by open access. What we have had to do is generate grass roots media coverage so that when people look for open access our name comes up in relevant channels throughout disciplines (#11). A few of those publications are the Times Higher Education, the Chronicle of Higher Education, Research Fortnight, and The Guardian.

Going back to the history of the project, it was on the sixth of January this year that I put up a temporary website asking for participants to get in touch with me. Within two hours, I have had over a hundred emails from people who wanted to be part of it. I told them that I wanted to build something like PLOS for the humanities, which meant a trans-disciplinary mega journal. We would cover all humanities disciplines and we would aim for swift publication in an open access format with the aim of being able to waive article processing charges that the author could not pay. It is all very well for people to say they want to help, but the true measure of whether a project is going to get off the ground is whether it actually has academic material at hand. To that end, we asked for pledges to publish. We needed to be able to go to funding bodies and libraries and say, "We have these people who have said they will give us material." We have 100 articles pre-pledged by academics for our first year. It is all very well for the people who come forward and volunteer to give you their articles, but they were the people who were keen in the first place. They are the digital generation who live on Twitter and are ready to publish their material openly anyway. What we also need to secure are publications from high profile professors at the top of their field in each discipline. To that end, we are trying to solicit at least 100 articles from those field leaders to cement the reputation from the top down.

In many ways I am trying to replicate a conservative environment within a very radical change in publishing practice. This results in what we are calling a Phased Introduction of Innovation. We have to start by looking identical to general publishing as it already stands. I wanted to point out when we think things will be happening (#13). For example, you can see in 2015 "Opt-In Changes to Peer Review." That would be for earlier adopters who are keen to specify that they would like their work to be evaluated in a post-publication mode of peer review. We would then have an evaluation phase over the course of a year and a half where we could collect actual evidence as to whether this diminishes quality and causes problems for research, rather than just speculation. There is fear, uncertainty, and doubt within certain parties that circulate about the viability of the method. In this way, I think we get the best of both worlds: the conservatives and traditionalists like us because they do not see any damage to what they perceive as a system that has historically worked to preserve quality, while the radicals see the potential for a transitioning change. We think that actually we managed to do the middle course because it is evidence based. We can try things out with the radicals, get the evidence, and then show it to the traditionalists who need persuasion.

Finally, I want to note that articles form a less important part of the ecosystem in the humanities than academic monographs of books. We are predominantly working with journal articles, but we recognize that books cannot be ignored. What we also recognize is that editorial labor, when it pertains to books, is substantially different from publisher labor in the case of journals. Time and time again when we speak to academics they tell us that they really value the editorial inputs into the writing of their books that can only be gained by publishers spending a career's worth helping with that process, shaping it, and knowing what makes a good academic book. That external input is very different to the process of publishing a journal article. To that end, we do not feel that we are at present the best placed to be taking on the publication of monographs. Therefore, we have chosen to partner with four presses who will undertake the editorial labor for us with financial cross subsidization from our enterprises. I cannot name the presses because agreements are in progress, but we have tentative agreements from three of the biggest university presses in the world. This is a mutually beneficial arrangement. For the presses, they get to experiment with open access at potentially no financial cost. They also get to distance themselves from a greedy for-profit model and into a mode that engages with scholar-run initiatives. Meanwhile, we get to associate with their prestige and brand name, which has substantial benefits in our need to persuade authors to submit journal articles. There is a cross subsidization of finance from journals to books in terms of money, and a cross subsidization from books to journals in terms of prestige and reputational capital. The provisional terms of that agreement include open licensing. We accept a range of open licenses. They will be free of charge at source the books in XML, HTML, and PDF formats while the publisher retains the rights to sell hard copies and other digital formats if they so wish. It is a provisional costing study that is the main aim of what we are doing here. It runs over five years, and the aim is to gather evidence as to whether this model of cross subsidization, if scaled up, could be a viable way for university presses to continue their valuable work while also ensuring that the serials ecosystem is completely free.

I will now move on to technical solutions (#15). There are some reasons for drilling down into the technology, because sometimes it is so entwined with the social solutions. One of the largest costs in the academic production process is typesetting. You might think this is redundant in the digital age, but it is not. We have to encode works into open formats so that they can be forward migrated for digital preservation purposes. As a quick example, there are certain formats that were available 10 years ago an Apple Macintosh that you cannot open today. If we work only in closed formats, we run that risk with scholarly articles as well, so we have to encode in this open XML format. Commercially available software for the typesetting process cost in the region of UK £20,000 per year. That is a lot of money. We have been developing at a fairly advanced stage an open tool to perform this process. It converts from Microsoft Word documents into an intermediary text encoding initiative format and finally into XML. This is being integrated into PKP’s open journal system software for future versions.

The next part of the technical process that is important is citation parsing. When you submit an article that has a bibliography, the computer is unable to know which articles are being referenced. It has to be broken down into constituent parts: author, title, journal, and only then can we know which articles are citing others. Essentially, we have got different levels of parsing. The reason that this is important is for moving away from the journal level as the metric by which we evaluate papers. We can only move towards the author or article level metric if we have ways of working out who is using material. Citation analysis continues to be one of the key metrics we have for that measure, even if it is flawed. We need for the technology to know who is citing who if we want to move away from journal brands towards the author.

The next topic is one of the most important technical aspects (#18). We have the notion of a mega-journal such us PLOS ONE, which is an interdisciplinary pot into which things are put when they meet the criterion of technical soundness. PLOS ONE does not focus on the novelty of science. It asks, "Is the science right? If so, it can go into this journal." This may be a problem for the humanities who still focus on the idea of niche journals that have histories of prestige. What we are proposing is to create system derived from arXiv, which is a US initiative with a system of overlay journals. I will illustrate this with an example. I edit a small niche journal at the moment called *Orbit* on the subject of Thomas Pynchon. I am able to do this because I am recognized in that field, and people are willing to say that, because I sanctioned it, it might have some value. They value the curation function that comes from my editorship. Once you see the editorial function as one of curating, it becomes clear that there is no reason why a journal edited by faculty who have reputation cannot be overlaid on top of a mega-journal structure. For instance, there could be a low barrier to entry for basic publication in the mega-journal, but when we come to the overlay journal like the one I run, it is an overlay on a mega-journal where every three months I am saying these are the 10 best articles that were published. It is a way of overlaying traditional systems of prestige conferred by academics on top of a new way of finding and publishing articles. Surprisingly, academics in the humanities like this idea. I was convinced it might be too radical for them, but actually they wanted it. We have already had at least five journals ask if they could transfer over to the system even before we are off the ground. This is a good start.

There might be other uses for this curation of issues in various levels of privacy settings. For example, if you are producing a course pack for students and couple it with print on demand, you could curate a course pack. The students get the link, they print on demand, it is with them five days later for a reasonable cost, or they can it read online if they do not want to pay at all. There are benefits and it could be used for a private research project. You could curate a selection of articles and get it printed.

I am going to talk briefly about

Digital Preservation. One of the main concerns that academics in the humanities have is about the ephemerality of material on the internet. They seem to forget that we have had to build vast structures for preserving material objects called libraries, but they think that things online stand a greater risk of disappearing. We have to come up with strategies to counter that perception. We have several systems for digital preservation and we have been extremely keen to stress that from the start to assuage the fear of those in the humanities that their work will disappear if I were to get run over by a bus. These systems are called LOCKSS, which stands for Lots of Copies Keep Stuff Safe, and CLOCKSS, Controlled Lots of Copies Keep Stuff Safe, plus off-site backups. We have also got the idea of a community maintained BitTorrent tracker for distributed peer-to-peer dissemination of articles to mobilize and engage the technical community.

The important part now is finance. How does this all get done? As we heard in an earlier talk, when labor is done, and there are systems, it needs to be compensated. If you cannot sell research as a commodity because it is open access, you have to find alternative ways of compensating the necessary labor involved in publishing. We have two phases of operation plans. The first is a philanthropically funded phase for five years. We have been in discussion with a major US funder about this and those discussions are ongoing. The second is a move to fully-fledged self-sustainability. By our estimation, the absolute upper (and ideal) cost of running this enterprise for five years is US $2.6 million, and that includes the monograph subsidy project. This would enable us to show that what we are doing is viable before we move to the model involving libraries paying us money. This is an evidence-based approach. If we can show it works, people are more likely to invest and think it is safe.

I have been stressing that the social problems are harder to solve throughout, so I put up a list of proposed staff members (#23). You will note that it includes no technical staff at all. The reason for that is we are planning to partner with an organization called Ubiquity Press in the UK. They have developed many of the technologies that we need already, and every time we contribute a piece of software back to them, they can lower the charge that they require per article. They are a non-profit enterprise that shares our goals, and we are dividing technical and social labor between them. We are employing staff primarily to fix social problems.

Last, but not least, is the sustainability model. I would like to propose that the problem with what we are doing at the moment is that we have a large number of libraries all paying a large amount so that they can privately rent content. What I propose is that we invert the subscription model. At the moment, many libraries are paying individually. Why do we not instead opt for a mode of collective procurement where a large number of libraries pay a small amount so that everybody can benefit from open access? In this way, all the libraries are paying for is infrastructure and labor rather than paying for the idea of an article as an individual commodity, which is what you get in the APC set up. It does not make sense to pay for services per article when you are paying for things to be open. Let us pay for the entire infrastructure and labor needed to sustain that kind of organization so that we can have an ongoing model of openness. If we obtain funding and we have a five-year run, it will be in year four that we will start to ask libraries to pay because by that time we would have shown our value. We would be asking approximately a thousand libraries worldwide, which is an optimistic estimate, to contribute a banded average rate of US$600. This is sometimes less than the cost of a single small niche prestigious traditional journal. We are not asking for much money and it is easily within the affordability of most libraries. It is also worth noting that this gets around the problem of double dipping. It looks to all intents and purposes like a traditional subscription.

質疑応答

●Q1 Thank you very much for coming to such a distant place and for making such a wonderful presentation. I have a question about the internationalization of the steering committee. So far, it seems like an Anglo-centric internationalization. Do you have a strategy to internationalize this committee? As far as I know, Dr. Melissa Terras of the Alliance of Digital Humanities organization has had relative success in internationalizing their academic committee.

●Eve I think it is a key issue that we need to address. We have an internalization committee and I would suggest looking on the website for that full list. We have members of UNESCO on that committee. We have also many representatives from China, but Japan is currently massively under-represented. If you are interested, please send me an email. From an Anglo-centric perspective it is very difficult to reach out into unfamiliar communities. That is our fault. I fully admit that. That being said, I am here now and it would be great to hear from people who would like to get involved.

●Q2　国会図書館の菊池です。質問は二つです。私は以前、国会図書館のメールマガジンで OLH のことを記事にしたことがあるのですが、そのときに分からなかったことをお伺いしたいというのが一つです。その内容は、ジャーナルの分野というものは、結局オールディシプリンになるのか、それとも個別ごとの分野になるのかという点です。関連して、実際の論文の刊行が始まる時期はもう具体的に見えているのかどうかもお伺いしたいです。

　もう一つが、先ほどの国際化に関連する質問です

が、国際化とは、論文の言語が英語でメンバーが国際的に参加するのか、それとも、メンバーが国際的になって論文の言語も各国語で行うのかという点をお伺いしたいです。

●Eve First of all, we are looking across all disciplines. We would have some kind of disciplinary taxonomy for submission, because otherwise it becomes impossible to conduct effective review. In regards to the second question, when I say internationalization, I mean the full internationalization of papers in languages other than English. I do not think it is acceptable for English to be the dominant cultural forum worldwide. It should be possible for those who do not want to write in English be able to do so. We could have simultaneous publication in English and in other languages, if researchers wanted that. We can have community translation or professional translation. Finally, in terms of timetable, I am very wary of nailing down specific dates because I will just set myself up for failure. On the other hand, I will say that next year we are planning to solicit papers and get the initial batch reviewed. We can only launch when the financing is firmly in place. If the financing comes through, we will be launching as soon as we can after the review of those papers.

●Q3 科学技術・学術政策研究所の林と申します。SPARC の運営委員も仰せつかっています。

今回のお話を聞いていて非常に印象的だったのが、プレゼンテーションの中に「パブリッシャー」や「ジャーナル」という表現がほとんど見受けられなかったことで、代わりに「オープンライブラリー」というネーミングを取られているところを勘繰りたくなるのです。そこにMartin さんはパブリッシャーになりたいわけでも、ジャーナルになりたいわけでもなく、ヒューマニティーを専門とする研究者のための、何かしら次の新しいメディアをつくりたいという思いが入っていると見るのは、うがった見方すぎるでしょうか。それがまず一つとは言いながら、実際に刊行されるものは、やはりPDF で毎月あるいは毎週発行されるジャーナルのようなものになるのか。その二つの点をお伺いしたいです。

●Eve Those are both very good questions. On the name basis it is worth stating that we are a publisher. This is going to be something that publishes frequently. We would be publishing on a rolling basis, not monthly or based on issues. When something is reviewed and is ready, it comes out straight away. To go back to the naming point, I think we chose the name “library” because the role of the library is changing in this century. Libraries do not hoard collections anymore. They are facilitators of the digital. To me that has to mean that it is bidirectional. The libraries are about helping researchers put material out as much as they are about helping researchers find material. We did think quite a lot about the name and we thought “library” encapsulates that bidirectional basis.