BIROn - Birkbeck Institutional Research Online

O'Sullivan, James and Eve, Martin Paul (2023) C21 Editions Interview with Martin Paul Eve. C21 Editions ,

Interviewer: James O'Sullivan   Interviewee: Martin Paul Eve

MARTIN PAUL EVE is the Professor of Literature, Technology and Publishing at the University of London's Birkbeck College. He is a founder of the Open Library of the Humanities and the Janeway publishing platform. All views expressed here are voiced as his personal opinion pertaining to his domain of academic expertise and do not reflect those of any employer.

September 28th, 2021

JOS: Digital editing and digital publishing, what do these terms mean to you?

MPE: Oh, I think they come with very different connotations to me. Editing has very much an implication of critical editions, a tradition of scholarly textual obsessiveness and pinning together manuscript histories, whereas publishing has a much broader frame in the digital world, it can include those digital editions but it includes everything from conventional, "just put a PDF online" through to more radical, interactive digital experiments, eLiteratures, hypertext and so forth. It covers academic publishing though to fictional publishing whereas the term editing to me carries a much narrower remit even in the digital world.

JOS: Focusing more on digital publishing, what do we do well, what might be better?

MPE: Ok that's a tricky question because ever since we've had digital publishing, people have theorised its possibilities and thought about hypothetical scenarios under which we could do things differently. So recent research I've done showed the initial inventor of the PDF, John Warnock, was told by Adobe to cancel the development of PDF because they didn't think there was any market for people wanting to replicate pages between different computer systems. There's all this talk in early 1993 desktop publishing about how we need to make publishing fit the medium in which it's created rather than just transferring print properties onto the digital world but the thing is that when people go to buy eBooks or they go to use scholarly published artifacts, they actually tend to want to go for the PDF nonetheless because various media constraints in use formats like citation principles or like the very fact they're just used to using pages draw them back to conventional print ways of working even though it's a digital space. So if you want to know what we do well and what we do badly I guess you have to measure that against how people behave and what they actually want from that system. It seems to me that actually, you know, for better or worse what many people want from a digital publishing ecology is simply the affordances of print publishing just made a little bit easier because you can download it and to that extent, we do that very well. If you're one of the more radical people who really wants to think about non-linearity of text, about hypertext possibilities, interactivity, embedded media formats and so on then various forms of publishing do that but it's not as widespread and normalised as it might be. So, I think you've got different demographics wanting different things and 'digital publishing' in quotation marks serves them differently in different contexts.

JOS: Do you think we'll ever get to a point wherein the digital is about more than just replicating print?

MPE: I mean the first thing to say is that there are various affordances of the print codex that are extremely good, you know, it remains an excellent mode of both sequential and random access compared to any eReader experience at the moment that will either let you do one or the other quite well. The find function is great in a digital text but still leaping between pages or even locations is quite a tricky task. I'm interested in the development of new eReader, eInk technologies, though. Like the ReMarkable for instance, which at once seemed to completely normalise that physical print culture – it bills itself as a paper alternative and really does feel and act like a piece of material paper – but on the other hand it really makes

the digital publishing environment one that works for me and I, in the last year to year and a half, with that device, have switched almost all of my reading, despite the bookcase behind me, to downloaded digital artifacts. It's worth saying also that I think the hardware technologies really do play a role here. There are lots of studies of how people read on backlit screens and what web styles of reading have done to our ability to process longform text on a screen, and the eInk technologies do not have that same change in neurocognitive dynamics of reading, as far as I'm aware, so really there is quite an interrelation between hardware, software, affordances, materiality or otherwise of these digital artifacts and that affects their uptake in the general population. I should say, though, the economics of that are complex because the ReMarkable is a *remarkably* expensive device that is not going to have huge mass market appeal; it is limited to an elite set of people who can afford that and who have reading as part of their professional activities.

JOS: And what about scholarship?

MPE: Well so, most of our journal publishing is now digitally "native" to use a term that most people don't like; most people are now used to using journals in a very different way to the conventional serial formats. I don't know many people who read paper format of journals cover to cover any more, we've a much more random style of access to that material and we dip in and pick out the articles we want. However, journals are not the only fruit and the long form publication of the book still remains key to many disciplines and there's a challenge in the scholarly space in that the scarcity of print correlates well with claimed proxies for accreditation. So, we have very few jobs in the academy. If there's a material investment in a press in printing your book it signals some kind of scarcity function that correlates with the kind of scarcities the evaluation panels are going through and so print form holds still value for those accreditation purposes. That transfers over into the digital publishing space because we want uniformity of cited objects and so we get trapped into replicating print, especially in the book space online. Despite attempts by Mellon Foundation to get university presses to do some experiments, people actually want their book to look like a book in the digital space and I think it's mostly for those reasons that I have just signalled: the prestige of print scarcity ecology.

JOS: What motivated the foundation of the Open Library of the Humanities?

MPE: Really a frustration that almost all Natural Scientific research will soon be freely available to read while almost all Humanistic research remains behind a paywall. I can't see that being a great environment for the future of Humanities or scholarship where we're expensive and science is free. It seems to me we have an increasingly educated population who enjoys studying the humanities at University and who would continue to engage critically with that type of material but without access through library structures we were stopping that kind of engagement that would be good for our disciplines. The thing about OLH is that, really, it's actually an economic model for pooling resources in order to avoid article processing charges in our disciplines, which just don't scale or work; they concentrate costs unaffordably. So, while we publish this is actually just a model like arXiv or Knowledge Unlatched of collectivising funding in order to provide open dissemination.

JOS: Do you think that open access should be considered an essential pre-condition of future publishing?

MPE: I think it is the logical conclusion of digital publishing in the scholarly space. This is complexly linked to systems of remuneration for the type of artifact that is produced, so there's a benefit in my mind to the fact that academics don't have to sell their work. If it were the case that every piece of scholarly published work had to recover its costs through sales, we'd see a rush to the bottom of people writing populist histories and trade books that are just all about celebrating Winston Churchill or whatever and to my mind what we have is

a freedom from that market in being able to research niche subjects that have a merit in scholarly interest but don't have a popular market interest. In that situation you're not remunerating people by sales so therefore they are to some extent liberated to give that work away so long as they have another channel that pays them to produce that work and that's where a university salary, increasingly hard to come by, enters the frame. So, the challenge is that on the one hand you now have this digital environment where people can give things away because they're supposedly paid and that's good for education it's good for not being trapped by the market and so on it's good for everybody's access. The challenges are that at the same time the number of university jobs is contracting, the contract where people used to be given a job for life at a reasonable but not excessive salary and a decent pension and security is being withdrawn, there's precarity among post PhD students, for instance. It's there really that we have a situation emerging, just as we get digital potential, where we lose the economic foundation that allowed people to give their work away within that structure. And so, I think it's not just a matter of what the digital does to the dissemination of products it's a matter of how people are remunerated and what that frees them to do. And if you break any part of that chain, you break the logic that says well these people are paid to create stuff they can give away for free digitally. If you're not paying them then they're not empowered to do that.

JOS: Is it possible that the affordances and ideals of open digital publishing can't be realised simply because of broader socio-economic considerations?

MPE: Absolutely. I mean you see this everywhere. You can look at the model Subscribed To Open that's come about in recent years which is a model for open access publishing. It's similar to OLH except the threat is that if not enough people participate or if libraries drop out, publication goes back to being toll access. The free-rider problem is the most theorised economic downfall for this type of initiative where if you require someone to pay for something that's free there will always be a number of people who get the thing for free without paying for it and if enough people do that, then your economic model collapses. Now in scholarly publishing the market of people who are willing to pay is quite a strange one: it's libraries who have been advocating for open access for decades so they're actually an audience who understand that and many of them do participate in these types of schemes but there's no universal guarantee that if their budgets are slashed they'll continue and, at the end of the day, if you've got to pay people for their labour, making something rivalrous and paid for is the best way to enforce that people have to pay for it. So yeah, it's quite a precarious balance here of different goods, different socio-economic situations and even if the technical affordances are what let us do the end thing in digital publishing, they are *not* the thing that conditions whether it is economically viable for people to participate in those affordances.

JOS: Do you think we've worked out models, both economic and technical, that can sufficiently sustain publishing into the future, both for large scale, institutionally supported projects and smaller independent projects?

MPE: Do you mean in the scholarly space or in…?

JOS: Across the entire spectrum; obviously, there are larger projects with institutional support, but then you also have grassroots publishers who may want to make materials openly available. Do you think that there are models that can scale to publishing projects of different sorts?

MPE: There are a range of models. I'd say that the primary challenge is that implementing them takes a lot of work and upfront investment. So people look at the OLH model and say 'oh you've got 310 libraries to support you' or whatever and that's great, you know, 'easy we'll do that too' but what they don't realise is that, you know, I was giving a talk in Japan one

day and then flying to America the next and then England the day after and then I had a stroke implementing this and getting it off the ground. The work in building this is huge, it's not 'build it and they will come', it's 'build it and then spend years telling people about it and persuading them of its merits and then maybe they will join'. So, the size of organisation that can afford to implement new models for digital publishing that are not well embedded and understood is a key factor. If you've got the resources to employ marketing staff to do it, yes you can. If you're a small grassroots independent publisher with just a few employees, not high throughput, I'd say it's much, much harder. I don't know whether that answers all of the questions.

JOS: Tell me…

MPE: I was going to say also though that there's a weird ecology at work at the moment where we still have a lot of subscription journals in the academic space... but every single academic journal article is available illegally to read for free through Sci-Hub. I don't know what is going on with the economics there but it's an interesting point: well libraries are still subscribing because they don't feel that Sci-Hub is a legal thing they can fall back on, which they are correct about, but it's interesting that the stuff is available for free even when it's being sold under DRM'd conditions for rivalrous access.

JOS: And is your suspicion that academics are giving work away themselves? Is this the attention economy at play?

MPE: Sci-Hub is much larger scale than that, it's a proxying download organisation that has a load of academic credentials from different institutions around the world and when someone requests an article it cycles through those credentials, tries to find one that has access then stores a copy of it in Library Genesis. So there is some of the attention ecology in how academics put their own stuff out there, as evidenced by ResearchGate and the recent number of takedown requests they got from Elsevier. But Sci-hub is a much larger scale, more professionalised outfit I'd say.

JOS: Tell me about Janeway. Why did you make that transition? What does it do that OJS doesn't?

MPE: Really the motivation was that, while I have huge respect for OJS and PKP, their code base is two decades old, it's written in PHP and we got to the point where we couldn't easily add features that we wanted for our own platform. So this was, again, a socio-economic decision: it's very easy to hire Python developers at a reasonable price. PHP developers are more senior, more expensive people and also to get to grips with a two-decade code base that has mostly its own libraries rather than external dependencies just meant that whenever we went to change something we were breaking it, we couldn't go through the usual upgrade cycle and we reckoned we could just do it ourselves in Python much quicker. (Laughs)

JOS: Do you think we're in danger of platform saturation? Wouldn't it be better if everyone was using a standardised tool or platform?

MPE: So interestingly there's a platform called the Next Generation Library Publishing Project that is looking at attempts to standardise some of the front ends of different platforms into a homogenised delivery interface. I'm not sure. They claim there's a market for this but we've heard different things that actually platform diversity is important and the way that platforms do things differently is valued by different organisations. And I guess this boils down to questions of centralisation versus decentralised competition in markets almost. What do you think is the most effective? If one tool does something really brilliantly and well then yeah, we wouldn't need everything else. But there isn't actually a consensus, well, what the proper workflow for scholarly publishing should look like in the future. Some people think preprints are really crucial and the future and the workflow should be just

preprint, post publication review and leave it in the preprint server. Others want a journal certification attached to that process. Others think preprint is the spawn of the devil and we shouldn't have it. And each different platform support those in various different ways. So I think, alright, you could call it saturation or just proliferation. I think generally speaking having more than one option is probably a good idea given that we don't know what problem wholly we're trying to solve in every domain.

JOS: What other tools and platforms do you find interesting?

MPE: The one I keep a close eye on is F1000's platform which does a really great job of post-publication peer review, of signalling ambiguities in researcher's feelings about the stuff their reviewing; so this is kind of a 'not quite sure about this yet' type mode. It's got really strong versioning and the preprint version that goes up is also incredibly well formatted because they have a strict process of ingest where you have to format your paper using their prefabricated tool. So that's a really slick platform, it's also closed source and owned by Taylor & Francis (they're Informa Group now) but they've been the preferred tender recipient for several huge projects like the Open Research Europe, the Gates platform and the Wellcome Trust's platform. So, you know, although there is platform proliferation, there is also a centralisation around what F1000 is doing and they have managed to consolidate most of the paid, big paid market for those huge funders that are part of the Plan S initiative. I always keep an eye on Elsevier and their purchase of Bepress in recent years was a really interesting acquisition. It caused basically a fracture in much of the library publishing world who thought that they were buying into a mission-oriented community led type platform and then found themselves in the hands of their 'mortal enemy' so to speak. So that for me is a really instructive case study in how values and your mission play a role in platform governance and sustainability and knowing that certain audiences in this space, the scholarly publishing space in particular have particular requirements of ethical ownership and ensuring that the organisations that are running their infrastructures share their values which you perhaps wouldn't get in a more commercially oriented publishing modality.

JOS: Magic wand, what would Janeway do that it can't do now? What does the ideal platform look like to you?

MPE: I think we don't do well enough the preprint workflow at the moment and this idea of post publication review. I'd really like to see more of that in my disciplinary spaces in particular but until we've got a platform that does it really well and makes people feel secure that the work will be fairly appraised and we've got a system for ensuring that you don't get abuse through that, it's quite hard. So we're still working on that side of things. There are several parts to the automation workflow later that I'd really like to improve around our final typesetting. I mean it's doable to go from JATS XML to a paginated HTML document using CSS Regions and then to print that as a paginated PDF. If we could get that working it would reduce our production costs per article by about £150 every go. Again, there's an interesting point there that though, to some extent one of our arguments has been, people have said 'what's the incentive on you to only publish good work this passed peer review?' and we say 'well, every time we publish something it costs us at least £150 plus all the staff time that goes into supervising that process'. If you eradicate that cost expenditure, does it somehow devalue what we're doing? Is it actually good that we can say that there is a cost per article that isn't just automated out of existence to some ways? I'm not sure that's quite right but there's a devil's advocate argument about economies of scarcity and value there that's quite fun to play with. Those are my two big things for Janeway at the moment. Versioning that's the last thing I'd say, we're not great at having versions of articles and redirecting people to say the most recent version and letting people revise things and upload fresh versions because again every time we do a new version, we incur that full typesetting

	cost. It's not easy for us to just let people say 'oh I want to update that now' and just link a new one on top.
JOS:	Why not adapt existing platforms could fit such purposes; I'm thinking here about systems designed to support, say, federated wiki structures?
MPE:	It's possible. I mean, so an interesting case study is Carnegie Mellon University, who adapted Janeway for their Encyclopaedia of the History of Science and they use a standard HTML or XML flow rather than having PDFs of their articles, which means that they're kind of in control of their own production and that has a pretty strong versioning feature built in. It's not an open wiki as you suggested or anything but it is a kind of modded version of our software that they use to do that better that we had. I think to be honest we will see and improvement in this area in the next few years because it's a demand that's coming from different quarters but I haven't quite squared the economics of what we'd do at our own publishing outfit around it. You can imagine some form of versioned wiki history of this stuff that's fine but it's really just how you layer the economics of production on top of that. Yeah, if Taylor & Francis (for example) let me modify my article whenever I wanted when it was published, they would presumably have to charge me to do it. We want to avoid charging authors in our model so where would we be funding this potentially exponential into the future number of modifications that we're having to re-typeset every time?
JOS:	There is this real division between scholarly and commercial publishing. Can you imagine a future where that division is less pronounced?
MPE:	It's interesting that some disciplines have more of a history of writing popular works that are trade crossovers for sale. So, your typical Cambridge historian's ultimate dream is to write the best-selling history book that is in every Waterstones and has a general readership. So, it's not as though there aren't points of academic writing that already crossover, it's not as though that didn't exist already, crossover with the general public. I think that what's in common with commercial publishing and scholarly publishing though despite the superficial appearance that isn't the case, is that it's all driven behind the scenes by various economic constraint and scarcity and we like to pretend that isn't so much the case in the Academy, you know, we're able to publish things that have only very niche appeal and that don't have a broad readership, yes, but as I pointed out, that's because we set up whole economic systems to facilitate it. Whereas the situation with publishing fiction at the moment is one that relies on sales. Now, you know, China Miéville, for instance is one author who has proposed an alternative socialist mechanism by which the state would supposedly recompense authors of fiction for their works that could enable them to be given away for free. I don't think we're going to see any change to that system under current government principles here in the UK at least but there are people thinking around this and what the digital might do. But in order to get there they have to think about how you reconfigure the mechanisms of labour and remuneration that go into authorship and publishing behind the scenes.
JOS:	Could you imagine a future where the platforms being developed by the OLH are used by smaller, independent publishers?
MPE:	Absolutely. I mean, we have a scholarly workflow kind of determining what we do, you know, there is a review mechanism but the workflow is actually customisable and you can rip any stage that you want out of it. So it's perfectly possible to skip peer review and just have something that's uploaded go straight into the production phase and it's possible to limit who can upload that so you can just have people in a back office loading content in. It's all doable it's just that, I suppose, one of the ways we've always marketed it is as a scholarly communications platform and that then predetermines the people who come and

find it in the first place. So, you know, there's something about, you can make your platform do things that other people might want, but how do they discover that that platform might be something they could use and that it would fulfil their purpose? It takes somebody with a bit of a probably technical mindset to look at it and say 'yes, well, let's use that, that will do it' but most people are not minded like that. They see a piece of web software and they don't know what they might do to change it to work for their purposes.

JOS: What is the future of the OLH?

MPE: Well, I'd like it just to keep existing given the dire economic worldwide climate! We've actually just opened for new journal applications for titles that are subscription that wish to flip to being openly accessible. So, you know, I've basically discovered over the years, we're not, well I kind of knew this from the start, we're not going to solve whole crisis of open access to humanities by ourselves but what we can do is represent that kind of regulative threat to the big organisations that dominate our publishing space and that are not moving, in my mind, quickly enough to implement radical models for OA. So every time we take a journal away from a commercial publisher, yes, we save some libraries a small amount of money because they don't have to pay that publisher next year but really we get disproportionate media impact from doing that and we're seen as the good guys fighting; David and Goliath against the big bad massive commercial publishers, we're doing the right thing, we're on the right side of history, we've got a great economic model so, you know, it all adds up and builds towards the cause of advocating for this being a good idea and in turn editors at those big presses turn around to their presses and say 'well what are we doing to make something work like this?' So that's kind of how I see it, it's not that we're going to grow and grow and grow until we're the size of Elsevier, it's that we're going to keep inflicting relatively low-level damage that has disproportionate media impact that makes people realise that there might be better ways of doing it and we'll see a gradual transformation more and more in that scope, I think.

JOS: Do you think it's possible that Elsevier will eventually impose a more radical model of open access?

MPE: Well to some extent transformative agreements are that model already, where they enter into basically bulk deals with institutions so that there isn't a fee per article because the institution has already paid upfront. Now, the challenge with those is the scale of them and the hyperinflationary rates that are being charged means that those deals are likely to eat up all the budgets of institutions that participate leaving no scope for the smaller players to get in on the action. So actually, the danger is that Elsevier, Taylor & Francis, Informa, whatever, Wiley, coming up with a model like that that people subscribe to and leaving no space for bibliodiversity and a long tail of smaller publishers who might have their own radical models. So, you know, I think if you look at the academic book publishing space and what was submitted to say UK's REF, you've got five or so massive publishers at the top and then this incredibly long tail of hundreds of small publishers with very niche subject areas, subject specialism and so on that could well be worth preserving but who could completely be priced out if all of the money just goes to the big four or five.

JOS: The broader aim of the C21 Editions project is to investigate and advance the practices of scholarly digital editing and publishing by researching and potentially prototyping data standards for various types of content including born digital content. With that very broad statement and objective in mind, what do you think are the essential things that we should look at?

MPE: Ok. So, I guess one of the questions is what are the standards that are already out there? You know, TEI is got to be high on your list there, JATS XML is another, BITS XML for books, the PDF standard itself and I guess the question would be if you were going to

develop new standards, why is another standard needed? So, what's the missing gap in this? One of the gaps seems to me to be a lack of a decent WYSIWYG editor for most of these formats that will run in a browser. We can't let somebody just paste their Word document into a browser and it generate good XML for us. If we could that would be a huge step forward. But, you know, that's a technological gap that might be worth thinking about is these standards, they may work to represent the data format but do they actually work for implementation purposes? Why has nobody been able to write that crucial piece of software in a way that actually works? I guess also you might want to consider issues of standards proliferation here and whether you're going to get that XKCD scenario where there's sixteen standards 'what we need is one to unify them all!' and now there's seventeen standards. You know, until there's an identified need for a new standard, why are we creating new ones? I think you've got different audience groups who need to be asked what they want and what's missing from this. So, I don't know if you've got Chris Ohge on your list to speak to from School of Advanced Study at London, he's got a book out on scholarly digital editing based around a Melville case study. I think he'd be a good person to talk to about what in his practices is missing when he tries to create a digital edition. I guess, then, also it would be good to poll some very bog standard academics about what they do every day when the do their research in a digital space and they come to a journal or an online book, you know, what do they actually do and probably the most likely thing is they download the PDF and if that's the case, you need to find your educators who aren't doing that. But it would be interesting to know: what are the majority of academics or end users of digital publishing platforms actually doing? Are they just seeking print replication and why? And they'll probably say 'because I have to cite a page number' and everyone online is just going to the digital edition because they have to cite a page number, which is really annoying, but yeah it could be the case and it would be good to know if that's the case, because it's my suspicion. So, I guess another question for you is: do you have a set of suspicions or hypotheses around user behaviour that could be tested on a relatively wide sample set to get a picture of how people behave, why they behave, and therefore what standards are being used? But yeah, those are my initial thoughts.