

New models for open digital collections?

This article discusses the potential for new community-based funding models to support digitization and open access (OA) publishing of digital collections. Digital collections of archival material such as texts, images and moving images are an important complement to journals and books in the ecosystem of scholarly resources that researchers, teachers and learners use. However, institutions find them expensive to acquire from publishers or to digitize themselves. In the US, Reveal Digital (RD) has set up a 'library crowdfunding' programme based on a cost-recovery OA model. The article describes how Jisc has collaborated with RD to introduce the model to UK institutions through their 'Independent Voices' collection of 20th-century alternative press materials and, in doing so, explores the potential and challenges for developing a similar approach in the UK.

Introduction

Digital collections of primary source materials such as texts, images and moving images have become an integral part of the ecosystem of scholarly content that researchers, teachers and learners use in their academic activity. They provide an important complement to other more traditional resources such as journals and books, especially within the arts, humanities and social sciences.

There is plenty of evidence of the impact that such collections have on research as well as teaching and learning. *The Impact of Digital Collections*, which is the culmination of a number of impact studies that were conducted over the years,¹ concluded that 'digital collections have become fundamental to modern scholarship', noting a shift towards humanities data science and data-driven research and at the same time potential for growth of the digital humanities, in particular in teaching.² Recent surveys also found that 'digital collections were not only essential to scholars' ability to access materials, but they influenced multiple aspects of their research practices' and contributed to the development of new skills and collaborative networks.³

How can we ensure that digital collections continue to be available and widely accessible to scholars and students against the backdrop of a tougher economic climate?

The 'golden age' of digitization

Typically, digital collections are the product of grant-funded digitization activity undertaken by the academic and heritage sector or are made available to the academic community by publishers, at a cost.

The first decade or so of the 21st century saw the buzz and excitement of what might be termed the 'golden age' of UK digitization, with high levels of public funding and support available. The scale of national programmes such as the New Opportunities Fund ICT Content programme⁴ would be unthinkable today. The programme distributed National Lottery funds and demonstrated the UK government commitment to a vision for increased access to a wide range of open resources to support lifelong learning through digitization. It included a £50m digitization strand and almost 150 projects within museums, galleries, universities and community groups. Initiatives such as Europeana, which brings together digitized collections from the European cultural heritage sector,⁵ and later the Digital Public Library of America, which



PAOLA
MARCHIONNI

Head of digital
resources for
teaching, learning
and research
Jisc



PETER FINDLAY

Digital portfolio
manager
Jisc

'How can we ensure that digital collections continue to be available and widely accessible ...?'

45 aggregates millions of digitized items from North American libraries, museums and archives,⁶ were enabled by the large amount of digitized material within the cultural and heritage sector. Big commercial players such as Microsoft and Google also started their book digitization programmes in the early 2000s in the UK and abroad.⁷

Within the academic context, it was also during this earlier period that Jisc initiated a number of innovative programmes aimed at democratizing access to commercially available digital collections across the academic sector. Jisc Collections, the Jisc service that supports the procurement of digital content for higher education and research in the UK, negotiated national purchases of resources such as Early English Books Online and Eighteenth Century Collections Online, which would have a lasting impact on scholarly practice in the years to come.

In parallel, through large-scale digitization programmes, Jisc supported the creation of resources we now take for granted such as the British Library 19th-century newspapers and many other open access (OA) collections. As Jean Sykes noted, 'Back in 2004 suggesting a small number of very large digitization projects was a bold and ambitious thing to do. But it has proved to be a seminal turning point in the UK where mass digitization is concerned.'⁸

Changing times

In more recent years, however, the climate in which we operate has changed – and become a lot tougher. Digitization funding has reduced, making this a competitive and time-consuming activity for institutions to pursue, especially in the context of OA. On the other hand, the high cost of content is a concern for most academic libraries.⁹ This increasingly has the knock-on effect of inhibiting the ability that libraries have to purchase other resources such as digital archival collections, and that 'end of year' budget that tended to be carved out for this kind of material is quickly drying up.

While there is a sense in the community that digitization has perhaps plateaued in the last few years, academic libraries have been trying to grapple with what a more strategic and co-ordinated approach to digitization activity might look like for some time. Chris Pressler's briefing on digitization of special collections both within RLUK libraries and across Europe, for example, was created a few years ago to support discussions in the development of a new RLUK strategy.¹⁰ The RLUK strategy also highlights the need to ensure that their libraries' rich collections are discoverable and used by researchers and students to maximize their impact.¹¹

More broadly, consultation that Jisc has carried out with library leads has revealed the need to support this kind of content in new ways, not least because of the recognition that the sciences tend to benefit from large grants but libraries also have to support the humanities. Libraries play a critical role in supporting the delivery of humanities courses. As humanities budgets shrink, libraries find themselves at the sharp end of satisfying student demand for resources which the students expect to be available, when they are now paying significant fees for their humanities courses, and this puts considerable pressure on library resources.

'libraries have been trying to grapple with ... a more strategic and co-ordinated approach to digitization'

An alternative way forward?

In recent years, alternative funding models based on the notion of the crowd have started to make their way into the academic and cultural sector. Crowdfunding has been seen as a likely way to fund more risky or unconventional research,¹² and digitization of heritage and archival material has been the subject of some big and small crowdfunding projects, from Neil Armstrong's space suit¹³ to the Peter Mackay archive¹⁴ and WW1 records.¹⁵ Nesta, the UK innovation foundation, has also recently completed an Arts and Heritage Matched Crowdfunding Pilot project to explore the potential of crowdfunding in the arts and heritage sector.¹⁶

46 However, it is more strategic and community-based funding models such as the Open Library of Humanities¹⁷ for journals and Knowledge Unlatched¹⁸ for books that are having a wider impact on scholarly communication. So, when US-based Reveal Digital¹⁹ approached Jisc about their 'library crowdfunding' model for digitization of OA collections, we were keen to explore how such a model might work in the UK.

Reveal Digital – what it is?

Reveal Digital (RD) started in 2011 with its first collection, 'Independent Voices'²⁰ (IV) which brings together a vast array of 20th-century North American alternative press publications. RD's approach is based on a cost-recovery OA model. The costs associated with digitizing, copyright clearing and delivering the collection are made available publicly. Once enough library pledges are accumulated, and after a brief embargo period, the collection is published on OA.

'it is more strategic and community-based funding models ... that are having a wider impact'

Pledging institutions benefit from early access to the collection in advance of it going OA, MARC records and copies of digital files, if they have contributed content. But above all they benefit from contributing to the strategic decision-making process of what gets digitized and who ultimately benefits from it. OA is a big driver for RD and the contributing libraries to ensure that once digitized, content does not become 'out of bounds' for many libraries due to often unaffordable paywall fees. While it is the pledging libraries that enable the digitization of all material through their financial contribution, it is important to note that *all* libraries, as well as society at large, eventually benefit from the investment of the supporting institutions once the content is made available on OA.

RD sees itself as a hub and facilitator of library OA publishing programmes by bringing together special collections archives, rights owners and contributing institutions. After the success of IV, which raised over US\$1.7m from more than 130 libraries mainly in North America, RD is moving towards an 'investment fund' model to better align themselves with the strategic investments that libraries are making in OA initiatives and to create a more sustainable funding base for their operations. This model will involve libraries making a multi-year commitment to a common fund and providing strategic input into the selection and prioritization of which collections get digitized. The end result will be the same, the production of OA digital collections that support the humanities, with a focus on 20th-century material.

'RD is moving towards an 'investment fund' model'

While in the UK we simply do not have the same scale or budgets as US universities, the emerging needs of academic libraries and scholars in relation to this type of resources are common. Might a similar approach work in the UK context?

Reveal Digital–Jisc collaboration

We felt that RD's approach for a sustainable OA model aligned strongly with Jisc's ethos. In addition, research undertaken by Jisc had highlighted academics' demand for digitization of 20th-century material, the 'black hole' of digitization, so IV provided a very good fit for an experiment. Back in September 2016, when we started discussing a collaboration with RD, only the University of Sussex in the UK had pledged for IV. After consulting with staff at Sussex Library on what drove their pledge – a mixture of academic demand and support for the OA model – we were encouraged that there might be wider interest in the community.

In order to make access to IV more affordable to UK institutions, we reached an agreement with RD based on Jisc-banded pledging fees.²¹ These were set as a one-off payment, with no recurrent annual fees, at a 65% discount on the US pledging rates, and a pledging period Jan–July 2017 (later extended to 31 December 2017). UK pledging libraries get the same benefits as US libraries in terms of early access to the full IV collection and MARC records. As an incentive to UK libraries, we also agreed that 50% of the contributions from UK institutions would be kept by Jisc and put towards a 'digitization fund' to digitize UK-based complementary material, where possible sourced from contributing libraries, to add to the

47 existing IV collection and also delivered in the UK by Jisc. Once digitized, the UK material will be available to everybody on OA. Participating institutions would set the priorities for content selection and digitization.

After holding an initial webinar in December 2016, followed by another one in June 2017, and some awareness raising through mailing lists and Jisc channels, it was clear that there were some enthusiastic early adopters and, at the time of writing (mid-September 2017), ten institutions had signed up: the universities of Sussex, Bristol, Sheffield, Manchester, the West of England, York, Reading, Salford, University College London and Birkbeck University of London. Pledging is open until 31 December 2017 through the Jisc Collections website.²²

'it was clear that there were some enthusiastic early adopters'

'Independent Voices UK' – new ways of working

Early conversations with the pledging institutions revealed that a key driver for them is developing OA collections. Even when an institution may not have any special collections of its own, they may have an active academic interest in content held by other participating libraries. By contributing to a jointly held fund, they support the opening up of content in perpetuity for the benefit not only of researchers and collaborative research, but also teachers, students and the public at large.

With Jisc agreeing to matchfund the small amount pledged by the libraries, in mid-July 2016 we started discussion with institutions on potential UK small/alternative press magazines to digitize as a complement to IV and how to approach the project as a whole. This in itself tests a new way for Jisc to work with its members. In the past, Jisc funded digitization programmes mainly with a view to enabling institutions to digitize their own collections. While Jisc had always been involved in developing standards and innovations in the way digitization is conducted and enabling exploitation of text and image and resource discovery, we had not directly co-ordinated selection and conversion activities.

As we approach the start of the project, it is clear that over the next year or so we will have to confront a number of challenges ranging from the nature of the 20th-century content we are dealing with, to approaches to selection criteria and copyright clearance, governance, operational workflows and final delivery and discovery of the digitized content.

'By contributing to a jointly held fund, they support the opening up of content in perpetuity'

Need for collaborative approaches

The material published by small presses is both narrow in one sense but immensely wide in another. Content may range from magazines (or zines) run off a mimeograph and cut and pasted on the kitchen table through to titles where production was undertaken along more traditional lines. Much of the material tends to be fairly rare as it is ephemeral in nature. Finding complete runs of a magazine is a challenge in itself. Quite often full sets are to be found in private collections. Institutions have often not collected this kind of material in a consistent manner so identifying complete runs will present a challenge.

An added complexity is that the content in question is 'owned' by communities associated with underground or small press publications. Many of these publications were driven by a particular personality or a group of personalities and are often related to a big cause such as ending racism, promoting collective action or challenging mainstream presses, so there are strong emotions at play for many of those who participated in making (and consuming) the publications.²³

48 The intention of the project is for each contributing institution to have a say in what is ultimately selected, and each institution may have differing expectations. There are therefore a number of stakeholders that need to be part of the process for making this content more accessible. It is not just a case of Jisc and participating institutions deciding what we wish to digitize. We will need to work as collaboratively as possible with rights holders and those who were involved in the creation of the content to understand their motivations and expectations.

Defining demand and selection criteria

Ideally, demonstrating sufficient academic demand for a certain type of content would be a key criterion to satisfy. However, in this case, reaching consensus will be challenging, as demand within institutions might differ. With 20th-century underground press publication, we are dealing with an emergent set of interests. Those studying modern history, cultural history, sociology, feminism and feminist thought, the politics of the counter-culture and works of avant-garde art and literature are some of the key audiences for these types of publication. But one issue is that there is no corpus (bounded set). These publications often have short runs, exemplifying particular political interests or reflecting the work of a particular group of poets or artists. This does not mean that they are any less important and in fact some of them had a significant impact on the course of publishing history.

'reaching consensus will be challenging'

We may find ourselves in a situation where demand is hard to define and where a 'long tail' of different kinds of publications are wanted by many, as typical within arts and humanities scholarship.

The copyright and moral rights challenge

OA, the primary driver for most of the participating institutions, implies use beyond a defined community, so ensuring that rights owners are on board will be paramount. Jisc will not necessarily have the resources to clear the rights so we will most likely engage a specialist to help us and resource that from within the project budget.

'ensuring that rights owners are on board will be paramount'

To enable the clearance in a reasonable time frame, we will seek out key rights owners but we will also explore possibilities afforded by the Orphan Works Directive (OWD). Identifying rights owners will be a great challenge, as demonstrated by previous experience such as the British Library's work on clearing permissions for the digitization of *Spare Rib*. Rights clearance was the biggest hurdle in that project with the team having identified over 4,500 contributors to *Spare Rib* and therefore potential copyright holders to clear permissions from.²⁴ In some instances the authors may be entirely anonymous. This presents an opportunity under OWD, but rights owners need to be acknowledged, even though their names may not be known. In some instances rights owners may wish to stay anonymous or in others, where they are identifiable, they may not want to be associated with something they produced in their teens or which was never meant for wider publication and distribution. Wherever possible, we will need to work with originators of publications so that they maintain a feeling of ownership. We will need to test if there are any scalable approaches or community frameworks to copyright clearance that we can implement for future activities, including the potential offered by the Extended Collective Licence.

The challenge of making content discoverable

With small magazines, we will most likely provide a series-based presentation of the content in the same way as RD have done with the US content. We see the same approach with how Jisc presented *Spare Rib*.²⁵ The grouping of the content is important in the same way that it is with archival material. With issue-based content on small runs, we are less dependent on platform-specific search functions, though full text can allow people to drill down to article level. However, with these publications, articles may be harder to define, the format of the magazine often being integral to how the content needs to be understood. If we are working with small magazines, we may be dealing with individual items, so we will need to work on

49 the specificities of search once we have made our selection. Small magazines often have novel formulations, such as being presented in a box or having strange sequencing, e.g. not volume one, two, three. A publisher might not have followed conventions, such as providing issue numbers, or may have used no sequential numbering, and this will present its own challenges in terms of presentation and the approach to discovery. We hope to ensure the content is discoverable by delivering it through existing Jisc content service as well as enabling cross-search with the IV collection.

'We will have to strike a balance between what is desirable and what is pragmatically doable'

Next steps

These are just some of the initial challenges that have surfaced as we begin work on this project in close collaboration with Jisc member institutions. Over the next few months as we progress our research on content, we will refine our approach to governance, selection criteria and clearing copyright and will devise a workflow for digitization and metadata creation. We will have to strike a balance between what is desirable and what is pragmatically doable and work in conjunction with the creators of the content, when possible.

We hope that this pilot project will provide us with a tangible output, and experience, for what can be achieved through a community-based approach to digitization and OA publishing. It will provide us with the starting point for a broader discussion with our member institutions, and other stakeholders such as the British Library and The National Archives, on what might be the elements and infrastructure of a model that can be adapted to the UK. We would be very interested in hearing the views of the readers. You can follow the course of the project and get in touch via the Jisc Content and Digitization blog.²⁶

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'Abbreviations and Acronyms' link at the top of the page it directs you to: <http://www.uksg.org/publications#aa>

Competing interests

The authors have declared no competing interests.

References

1. TIDSR: Toolkit for the Impact of Digitised Scholarly Resources: <http://microsites.oii.ox.ac.uk/tidsr/> (accessed 25 September 2017).
2. Meyer E T and Eccles K, *The Impact of Digital Collections: Early English Books Online & House of Commons Parliamentary Papers*, 2016, London, Jisc and ProQuest, p 52–53: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740299 (accessed 25 September 2017).
3. Green H and Courtney A, 2 February 2017, Digital collections offer researchers opportunities to develop new skills and scholarly communications networks, *LSE Impact Blog*: <http://blogs.lse.ac.uk/impactofsocialsciences/2017/02/02/digital-collections-offer-researchers-opportunities-to-develop-new-skills-and-scholarly-communications-networks/> (accessed 25 September 2017).
4. Education for Change, *The Fund's ICT Content Programmes. Final evaluation report*, 2006: http://www.biglotteryfund.org.uk/er_eval_ict_final_rep.pdf (accessed 25 September 2017).
5. Europeana: <http://www.europeana.eu/portal/en> (accessed 25 September 2017).
6. Digital Public Library of America: <https://dp.la/> (accessed 25 September 2017).
7. Tran M, Microsoft teams up with British Library to digitise books, *The Guardian*, 4 November 2005: <https://www.theguardian.com/technology/2005/nov/04/microsoft.books> (accessed 25 July 2017).
8. Sykes J, Large-scale digitization: the £22-million JISC programme and the role of libraries, *Serials*, 2008, 21(3), 167–173; DOI: <https://doi.org/10.1629/21167> (accessed 25 September 2017).
9. SCONUL Content Strategy Group: <https://www.sconul.ac.uk/page/content-strategy-group> (accessed 25 September 2017).
10. Pressler C, National Digitisation Review: Shifting Sands, *RLUK Briefing Paper*, 2014, Dublin City University, RLUK: <http://www.rluk.ac.uk/wp-content/uploads/2014/12/RLUK-National-Digitisation-Review-CPressler.pdf> (accessed 25 September 2017).
11. Nothing Hidden, Nothing Lost: Exposing and exploiting our collections, *RLUK strategy 2014–17*: <http://www.rluk.ac.uk/strategy-2014-17/hidden-lost-exposing-exploiting-collections/> (accessed 25 September 2017).

12. Murray J, Crowdfunding: the new buzzword for academics needing research money, *The Guardian*, 25 Oct 2016: <https://www.theguardian.com/education/2016/oct/25/crowdfunding-academic-research-money-brexitees-isd> (accessed 25 September 2017).
13. Reboot the Suit: Bring Back Neil Armstrong's Spacesuit: <https://www.kickstarter.com/projects/smithsonian/reboot-the-suit-bring-back-neil-armstrongs-spacesu> (accessed 25 September 2017).
14. Help Stirling to protect The Peter Mackay Archive: <http://www.crowdfunder.co.uk/petermackay/> (accessed 25 September 2017).
15. Lest we forget: <https://oxreach.hubbub.net/p/lestweforget/> (accessed 25 September 2017).
16. Nesta: Arts and Heritage Matched Crowdfunding Pilot: <http://www.nesta.org.uk/project/arts-and-heritage-matched-crowdfunding-pilot> (accessed 3 October 2017).
17. Open Library of Humanities: <https://www.openlibhums.org/> (accessed 25 September 2017).
18. Knowledge Unlatched: <http://www.knowledgeunlatched.org/> (accessed 25 September 2017).
19. Reveal Digital: <http://revealdigital.com/> (accessed 25 September 2017).
20. Independent Voices: <http://revealdigital.com/independent-voices/> (accessed 25 September 2017).
21. Reveal Digital Independent Voices Collection 2017–18 Agreement: <https://www.jisc-collections.ac.uk/Catalogue/Overview/index/2586> (accessed 25 September 2017).
22. Reveal Digital Independent Voices Collection 2017–18 Agreement, ref. 21.
23. McMillian J, *Smoking Typewriters: The Sixties Underground Press and the Rise of Alternative Media in America*, 2011, New York, Oxford University Press and Fountain N, *Underground: The London Alternative Press, 1966–74*, 1988, London, Routledge.
24. Vernon A, 28 May 2015, Digitising Spare Rib magazine: The inside story, *Living Knowledge blog*, *British Library*: <http://blogs.bl.uk/living-knowledge/2015/05/digitising-spare-rib-magazine-the-inside-story.html> (accessed 25 September 2017).
25. *Spare Rib*: <https://journalarchives.jisc.ac.uk/britishlibrary/sparerib> (accessed 25 September 2017).
26. Jisc Content and Digitisation blog: <https://digitisation.jiscinvolve.org/wp/> (accessed 25 September 2017).

Article copyright: © 2017 Paola Marchionni and Peter Findlay. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author: Paola Marchionni
 Head of digital resources for teaching, learning and research
 Jisc, GB
 E-mail: Paola.Marchionni@jisc.ac.uk

ORCID ID: <http://orcid.org/0000-0002-9544-5410>

To cite this article:

Marchionni P and Findlay P, New models for open digital collections?, *Insights*, 2017, 30(3), 44–50; DOI: <https://doi.org/10.1629/uksg.375>

Published by UKSG in association with Ubiquity Press on 08 November 2017