

Changing the face of scholarly information provision: a case study of developing and launching JISC eCollections

The primary role of JISC Collections is the licensing of content on behalf of its UK higher and further education member organizations. Affordable pricing and flexible terms and conditions of use, enabling content to be used to support teaching, learning and research, are central to negotiations. In 2011, JISC Collections launched JISC eCollections (www.jiscecollections.ac.uk), a community-owned content service developed to protect and preserve existing content investments by offering an alternative to commercial providers. It brings access fees within the control of the community, in terms of ring-fenced reinvestment and moderation of increases. A completely new venture for JISC Collections, the rationale behind the service's development and some of the challenges overcome in creating three new content platforms are presented in this case study. These included discovering and addressing gaps in content, rationalizing metadata, bringing lots of licences together under one model, anticipating and meeting the needs of multiple audiences (but delivering a simple overall interface) and developing an appropriate business model.

Background

JISC Collections has been licensing content in perpetuity for over ten years: the first agreement was signed in 2002, for ProQuest's Early English Books Online (EEBO). Since then, several more licences have been negotiated, for historic books, journal archives and multimedia content such as documentaries and educational films. In 2010, JISC Collections invested a further £2.5 million in film and image content representing UK and world history of the last 25 years, which was specially selected for teaching and learning. In total, JISC Collections (with institutional funding via JISC) has invested £20 million in centralized licensing of critical archives and content collections that had previously been prohibitively expensive for most institutions (or not even electronically available). Access to this content has typically been via the publisher or, in the case of the multimedia content, via platforms managed by UK data centre, EDINA, the costs of which were funded by JISC. Most of the journal archives can currently be accessed by JISC Collections members without further charge, but access charges are levied for the historical book collections, and the real costs of the multimedia platforms are hidden through JISC funding.

In this context, JISC Collections sought to protect and preserve existing and future investments by developing an independent service to ensure that future access fees for perpetually licensed content would, as far as possible, be unaffected by commercial levels of inflation. In this way, the education community could take ownership of its acquisitions and be assured of future control. Consolidating the range of licensed content would also simplify both the user experience and the administrative management of licensed content, and increase the ease with which it could be exploited for teaching and learning. The vision was to introduce economies of scale and thereby provide a sustainable, value-for-money alternative to licensing content on commercial providers' platforms, with a single access fee and entry point. The aim was to simplify the discovery and increase the usage of licensed material, particularly by undergraduates. Although developing a content provision service



CAREN MILLOY
Head of Projects
JISC Collections

75 represented a departure from the existing service offered by JISC Collections, its knowledge, skills and networks (and those of its project partners, the UK data centres EDINA and Mimas) meant it was well placed and well equipped to make the transition.

A Ford Escort, a Mini and a Ferrari – simplifying access and use

In 2009, JISC published a study called *User Behaviour in Resource Discovery (UBiRD)*, undertaken by Professor Wong and his colleagues at Middlesex University¹. The recommendations from this study were very influential in making the case to JISC to fund the development of three new platforms; the study highlighted the frustrations that users face in the usability of publishers' platforms and library systems and the blurring boundaries between information literacy and digital literacy. Many users were 'fluent in Google', but once within a publisher or aggregator platform, were confused by the differences between interfaces, which required users to 're-frame' their minds every time they moved from one platform to another. This finding chimed very closely with the results of both the JISC national e-books observatory project² (where students were simply unable to understand why some websites functioned in one way and others didn't) and the British Library/JISC *Information Behaviour of the Researcher of the Future* study³ (which overturns the common assumption that the 'Google Generation' – youngsters born or brought up in the internet age – are innately web-literate). The best way to describe this (credit to Professor Wong) is to think of cars: there are enough similarities in all cars to allow you to climb in and start driving, even if the build and features are different – you can recognize enough to know how to use it.

" This finding ... overturns the common assumption that the 'Google Generation' ... are innately web-literate."

It was therefore a core requirement to group the three different content types (historic books, journals and multimedia) into specially developed platforms with clean, simple interfaces that cater both to the Google-level searcher (typically undergraduates) and to the more sophisticated Boolean searcher (post-grads and researchers). Users would then have to visit and become familiar with only one platform to explore each type of content, rather than visiting each publisher's platform separately and learning how to 'drive' all over again. In addition, librarians only have to manage and link out to three platforms (rather than 14, and counting) and everyone has access to everything. Simple.

Except it's not simple. Not by far.

I trust that you received the package

In every agreement, JISC Collections has the right, on behalf of the UK education community, to locally host the content licensed for the long term. The publishers of the content provide a copy of the content on hard drives that are securely kept and preserved in case their use should become necessary in the future. Prior to development of JISC eCollections, the contents of most of these hard drives were not thoroughly checked on receipt; upon loading and indexing the data for inclusion in JISC eCollections, it became apparent that some content and metadata was missing. Developing the platforms and loading the content has therefore highlighted a weakness in previous processes, identified data gaps to be filled in accordance with existing licences, and demonstrated the need to verify that the hard drives (the asset of high value) do really contain all the data they are meant to in future. Common sense, but not a simple task until you have the method and the platform to do this.

The language of the law

JISC Collections uses a model licence for all its agreements. This is important because it helps to standardize the many terms and conditions of use, acts as stamp of approval and helps librarians communicate terms of use to users. In presenting the 'simple' philosophy to end users and librarians, a complex and long task started within JISC Collections. The aim was to have only two sub-licences for all the content – one for the JISC MediaHub platform and one to cover both JISC Historic Books and JISC Journal Archives. In addition, the sub-licence had to have some future proofing to allow for possible innovations such as

76 crowd-sourcing activities. To achieve this, it meant negotiating Variation agreements with all the publishers – 42 in total – and getting agreement to a revised sub-licence agreement without each publisher making amendments.

This was a major undertaking which took longer than anticipated, in part because some of the multimedia content owners were hard to track down, with contacts having moved on since the last round of variations in 2006, and some organizations now in receivership. If just one change to the sub-licence is made by one publisher, that change has to be acceptable to all the others and it can happen that the lowest common denominator is accepted. JISC Collections fought hard to keep the terms and conditions the same and to get the future-proofing clauses included. In all negotiations, lawyers like to make changes to language; while these are typically harmless, in our case such changes needed to be approved by all (or rejected in the first instance), so discussions took longer than usual on several points.

Some of the hardest clauses to negotiate were for text and data mining, open metadata and the creation of new metadata to supplement the metadata provided by the publisher. To support discovery, the development path taken by EDINA was to make metadata and thumbnails for all the content fully open and discoverable on the web. Agreeing these clauses meant explaining to providers, at some length, the benefits to educational users – and to the content owners themselves – of openness.

The end result is two standard sub-licences (one to cover JISC MediaHub and one to cover both JISC Historic Books and JISC Journal Archives), and common terms and conditions of use across all three platforms. A complex task, with hours of negotiation, has been worthwhile because of the benefit to librarians and users of simple, consistent licensing.

The meat of the issue: metadata

JISC Collections had the content and it had the licences. But what about the metadata? Metadata enables organization, discovery, use and understanding. In the Digital Images for Education project,⁴ where films and image collections were purchased through a tender process, a metadata schema was created and content owners had to provide their metadata within this. However, that did not always mean that the content of the metadata was accurate or relevant despite a validation process. For example, the key words entered into a specific descriptor field, such as 'man with book', may have been very useful to picture researchers at newspapers, but not necessarily to students at college and university. The picture itself was deemed relevant to educational courses by a group of educational evaluators, but the metadata itself, even in compliance with the schema, may have been irrelevant or insufficient.

EDINA had been working with multimedia metadata for many years, and in creating a single platform for JISC eCollections (to merge Film & Sound Online, NewsFilm Online and Education Image Gallery) had to bring the metadata for over 30 collections together onto one platform. The same action had to be performed for all the journal archives and the historic book collections by Mimas; although EDINA had been through processes of this nature before (for example, on projects such as Education Image Gallery and Digimap), it was a first for Mimas (and JISC Collections).

Rationalizing metadata provided in different schema and formats is a very complex task, and if not done correctly can have a major impact on discovery and search results. JISC Collections staff spent a long time looking at the metadata for EEBO, ECCO and the British Library collections of 19th Century books with library IT expert Owen Stephens. Each metadata schema was different, so a large spreadsheet was created to determine what fields needed to be migrated, searchable, displayed and filterable and a fun day populating the spreadsheet ensued.

"If just one change to the sub-licence is made by one publisher, that change has to be acceptable to all the others and it can happen that the lowest common denominator is accepted."

"Rationalizing metadata provided in different schema and formats is a very complex task ..."

77 As the day progressed, it became frustratingly clear that this process should have taken place at the very beginning, as EDINA had done for JISC MediaHub. Mimas and JISC Collections had not been able to do this, however, as comprehensive metadata for all texts within EEBO was not part of the original licence agreement due to cost. (The only metadata available for EEBO had been created as part of the TCP project, and covered just 20 percent of EEBO titles). JISC Collections explored other potential routes by which to license metadata to support EEBO, so that all the books would be discoverable during a search, but as the JISC Historic Books platform neared beta testing, it became clear that the only way to capture the metadata and ensure discoverability (and therefore usability) of the collection was to absorb the cost and purchase the MARC records from ProQuest. Therefore the metadata analysis took place quite late in the process, and while JISC Collections is not able to warrant that the metadata is of best quality, at least all the content on all three platforms can be found, viewed and filtered.

Improving the quality of the metadata for JISC Historic Books and JISC MediaHub is certainly a development for the future. The next challenge is making sure each platform is discoverable via library discovery systems (such as Primo, Summon and EBSCO's EDS); however, this will require interaction between the discovery systems providers.

Do you drive an automatic or a manual?

One of the aims of the JISC eCollections service is to encourage a broader range of institutions and users to make use of the content which has been licensed on their behalf, hence the focus on simplicity. Another aim is to aid the making of new discoveries and connections between the content collections on each platform. The default setting is for searches to be run across all content on the platform, to throw up new connections and perhaps results that would not have been found before; users can then opt to filter by a specific collection.

All the platforms were therefore developed to be 'simple' and 'wide' in accordance with the UBIRD study, and the three-column interface of JISC Journal Archives and the one-click filtering of JISC MediaHub have been well received. The single Google-style search box of JISC Historic Books, however, received mixed reviews, especially from academics used to using the ProQuest and Cengage interfaces.

A survey of the JISC Historic Books Advisory Board and interviews conducted by Owen Stephens for the metadata study quickly showed that while some will be happy with using a single search box and then filtering, many academics know exactly what they want and prefer to enter a bibliographic reference number or a particular imprint, for example, to find the title they want immediately. While this approach negates the serendipity, it is no less valid as a search methodology and therefore the platform was reconfigured to ensure that existing 'manual' searching techniques were catered for, but did not interfere with the simple design so that all users would recognize and be able to drive the platform.

Self-sustaining versus unfettered access

JISC Collections is funded by JISC, which in turn is funded by the UK HE and FE funding councils. The content licensed by JISC Collections in perpetuity is effectively owned by the education community.

Perhaps the most difficult challenge in developing the JISC eCollections service was in deciding the business model to support the ongoing costs and developments to the service. If the service were to be freely available to all UK institutions, access would be open and there would be no barriers, but it would rely on continued funding from JISC to support the costs of the service. In the current economic climate, sustainability is critical and it would be short-sighted and a potential risk to be reliant on one single source of funding. Consequently, the business model for JISC eCollections is a delicate balance between generating revenue to support the service's ongoing provision and development, and ensuring cost-effectiveness for members. The service fees cover all three platforms so that

78 all subscribers can access all content as part of their membership. The fee is as affordable as possible, to make it as easy as it can be for institutions to join the service, and the revenue received goes directly back into supporting the service. The long-term business plan for JISC eCollections, under which the service will become self-supporting, includes generating revenue from alternative sources; identifying how this might work was a major challenge for a new service.

Taking ownership and driving forward

The core vision of JISC eCollections is that it is a 'community-owned content service' – that the education community take ownership of their content and developments to the service. Advisory boards consisting of librarians, teaching staff and researchers have been set up for each platform. The remit of these boards is to discuss new opportunities and to make sure that future developments and content licensing support use in education and research, and contribute to the ongoing sustainability of the service. The boards are already exploring international collaborations to help improve the quality of the metadata and text transcriptions through crowd-sourcing as well as how new technologies may be used to support integration into the services and systems in which users typically start their content journey. Focus groups are planned to help inform the creation of digital literacy and teaching tools to link content with learning and help build usage of JISC eCollections into curricula, such as lecture synopses, slide decks, video lectures, reading lists, discussion guides, introductory essays and essay questions with associated research pathways. All these initiatives will be driven by the community to shape JISC eCollections and to help ensure that members have access to a unique collection of heritage content.

“The core vision of JISC eCollections is that it is a 'community-owned content service' ...”

In conclusion: significance of JISC eCollections

Simultaneously developing and launching a suite of new platforms for a diverse range of content has been a complex project; in a sense, it is fortunate that it was not possible to anticipate all the challenges ahead as this may have moderated JISC Collections' ambition. Instead, a service has been delivered that has the potential to change the balance of scholarly information provision in the UK and to enable the community to take ownership.

“ ... a service has been delivered that has the potential to change the balance of scholarly information provision in the UK and to enable the community to take ownership.”

References

1. JISC user behaviour observational study, User behaviour in resource discovery. William Wong, Hanna Stelmaszewska, Balbir Barn, Interaction Design Center; Nazlin Bhimani, Learning Resources, Middlesex University Business School; Sukhbinder Barn, Middlesex University, 2010, London, JISC. <http://www.jisc.ac.uk/publications/programmerelated/2010/ubirdfinalreport.aspx>
2. JISC national e-books observatory project, <http://observatory.jiscebooks.org/>
3. British Library / JISC, *Information Behaviour of the Researcher of the Future*. Peter Williams and Ian Rowlands, CIBER, University College London, 2008, London, British Library / JISC. <http://pressandpolicy.bl.uk/Press-Releases/Pioneering-research-shows-Google-Generation-is-a-myth-32b.aspx>
4. Digital Images for Education: <http://www.jisc-collections.ac.uk/Catalogue/FullDescription/Index/30> (accessed 6 Jan 2012)
5. Text Creation Partnership: <http://www.lib.umich.edu/tcp/ebo/description.html> <http://www.jisc-collections.ac.uk/Catalogue/FullDescription/index/1006> (accessed 6 Jan 2012)

The author thanks Charlie Rapple of TBI Communications for assistance in preparing this manuscript for publication.

Article © Caren Milloy

Caren Milloy, Head of Projects

JISC Collections

Tel: +44 (0)20 30066003 | e-mail: c.milloy@jisc-collections.ac.uk | Twitter: [carenmilloy](https://twitter.com/carenmilloy) | web: www.jisc-collections.ac.uk

To cite this article:

Milloy, C, Changing the face of scholarly information provision: a case study of developing and launching JISC eCollections, *Insights*, 2012, 25(1), 74–79, doi: [10.1629/2048-7754.25.1.74](https://doi.org/10.1629/2048-7754.25.1.74)

To link to this article:

<http://dx.doi.org/10.1629/2048-7754.25.1.74>